AD_____

GRANT NUMBER DAMD17-96-1-6288

TITLE: Reaching Rural Mammographers for Quality Improvement

PRINCIPAL INVESTIGATOR: Nicole Urban, ScD

CONTRACTING ORGANIZATION: Fred Hutchinson Cancer Research Center
Seattle, WA 98104-2092

REPORT DATE: October 1997

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

REPORT DOCUMENTATION PAGE

OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1997	Annual (16 Sep 96 - 15 Sep 97)

4. TITLE AND SUBTITLE
Reaching Rural Mammographers for Quality Improvement

5. FUNDING NUMBERS
DAMD17-96-1-6288

6. AUTHOR(S)
Nicole Urban, ScD

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Fred Hutchinson Cancer Research Center
Seattle, WA 98104-2092

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

The Fred Hutchinson Cancer Research Center, the University of Washington, and the Washington State Department of Health are collaborating to develop and implement a mammography quality improvement program (MQIP), in order to demonstrate its feasibility and effectiveness for dissemination. The MQIP emphasizes continuous quality improvement (CQI) in film interpretation, within the context of a comprehensive program designed to meet the requirements of the Mammography Quality Standards Act (MQSA) of 1994.

During the first year of funding, project investigators and staff focused on the development of procedures and materials necessary to recruit and conduct the CQI. Recruitment has begun for this part of the study. Additionally, planning and preparations have gone into developing training sessions for tumor registrars and mammography technologists. The Mammography Tumor Registry (MTR) database which will support the MQIP has also undergone structural changes to allow for the addition of MQIP facilities that do not use computerized systems. The database itself has also been modified to make the process of adding and validating data from MQIP facilities and report generation more efficient.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Mammography	93
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

__X__ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Nicole Urban 10/14/97

PI - Signature Date

Annual Report for Grant DAMD17-96-1-6288

September 16, 1996 - September 15, 1997
Year 01

Reaching Rural Mammographers for Quality Improvement

Nicole Urban, ScD
Principal Investigator

Table of Contents

INTRODUCTION

It is generally agreed that regular screening by mammography is a woman's best strategy for preventing death due to breast cancer. However, mammography quality is of concern for three reasons. First, recent evidence of variability in radiologists' interpretations of the same mammograms suggests that improvement is needed in mammographers' accuracy in reading films.[1] Second, growing attention to issues of costs and cost-effectiveness suggests the importance of improving specificity in reading mammograms.[2,3] Third, the efficacy of screening younger women remains controversial.[2,4,5,6,7]

The purpose of this project is to demonstrate that accuracy in film interpretation can be significantly improved by a continuous quality improvement (CQI) program which provides mammographers with immediate feedback on their interpretations of films specifically selected for their teaching value.

The Fred Hutchinson Cancer Research Center (FHCRC), the Department of Radiology at the University of Washington (UW), and the Washington State Cancer Registry (WSCR) at the Department of Health (DOH) are collaborating to develop and implement a Mammography Quality Improvement Program (MQIP), in order to demonstrate its feasibility and effectiveness for dissemination. The MQIP emphasizes improvement in film interpretation, within the context of a comprehensive program designed to meet the requirements of the Mammography Quality Standards Act (MQSA) of 1994.

The MQIP employs routine systematic monitoring of measurable outcomes of screening mammography, including sensitivity, specificity, and positive predictive value. This is referred to as its *surveillance* function. It also identifies for mammographers their false positive and false negative cases, so that they can improve quality through review of their own films. This is its *audit* function. In addition, it provides continuing education for radiologists, and training for technologists, as required by MQSA. This is its *certification* function. Most importantly, it incorporates immediate feedback following a radiologist's interpretation of practice films selected for their educational value. This is its *continuous quality improvement (CQI)* function. The MQIP is comprehensive, and will ensure that participating facilities are in compliance with evolving accreditation rules.

The primary objective of this project is to develop and implement a comprehensive, continuous MQIP in 16 mammography facilities located in rural areas of Washington State. The demonstration of the MQIP involves development and implementation of the 4 functions noted above.

A research study is being conducted within the demonstration project. The **primary research objective** is to determine if the CQI program can increase the accuracy with which mammographers interpret films. **Secondary research objectives** are to 1) determine inter-rater variability in film interpretation in a set of films selected for their teaching value, before and after implementation of the CQI program; 2) determine post-CQI intra-rater variability in film interpretation; 3) determine if digitized films can be interpreted with the same accuracy

as can high-quality copies of films; and 4) determine if the accuracy with which films are interpreted depends on covariates, the age of the woman being of particular interest. The availability of comparison films will also be considered as a covariate.

This four year project is currently at the end of its first year.

BODY

In order to achieve project goals, 18 major tasks needing to be accomplished were identified in the original statement of work. During year 01 of the project, major headway was made on several of these tasks.

The design of the CQI is the core of the MQIP. It is imperative that all aspects of the CQI be carefully constructed. To this end, a manuscript describing the design of the study and the analysis of results has been accepted for publication and is included as Appendix A.

To implement the CQI, research was first conducted to identify 16 mammography facilities meeting project criteria. Each facility must be located in a rural area and must not be associated with any other participating facility. In addition, facilities were selected so that their combination provided at least 60 radiologists who would be eligible for participation in the CQI.

Once facilities were identified, recruitment materials and procedures were developed to solicit the participation of these facilities and their radiologists in the project. Detailed recruitment procedures as well as materials used in the recruitment process are attached as Appendix B. The recruitment plan was begun in the latter half of the first year of funding and will continue into year 02. In addition, to ensure the participation of radiologists, CME credit for participation in the CQI was obtained through the University of Washington.

Additional work for the CQI portion of the MQIP has involved the development of software designed to present digitized mammograms presented on a personal computer. The software allows the radiologist to indicate the location of any potential malignancy, and select an assessment of the mammogram. The software tracks each assessment made by the radiologist, scores the entire test session, and provides feedback for incorrectly read mammograms. At the end of year 01, the software is 95% complete and in the debugging phase.

To perform the CQI, 180 mammogram studies meeting project criteria must be obtained, copied, and digitized. Mammogram studies that can be included in this study must be challenging, but interpretable. Approximately half will contain a malignancy. Specific procedures for selecting, obtaining, and handling films were developed during the first year and the process of selecting the films has begun. To obtain such a large amount of films meeting project criteria, it has been necessary to solicit the help of community radiologists. A great deal of time had been invested in contacting radiologists and assisting them with film selection. To date, approximately 60 studies have been identified.

Once all film studies are identified, digitized, and entered in to the CQI software, the process of conducting a CQI session will be piloted on a test facility. A mammography facility located in a rural area has been identified and recruited to participate as the pilot site. Currently, project staff are in the process of linking this facility to the Mammography Tumor Registry (MTR). The MTR was established three years ago as a registry of mammography data from local facilities linked to Washington State and SEER Cancer Registry data. The purpose of the MTR is to provide surveillance as well as feedback to participating facilities and radiologists.

To address the surveillance component of the MQIP, work has been done to modify the MTR to allow for the inclusion of facilities that do not use a computerized system to store their mammography data. A paper worksheet has been developed and will be presented to paper-based facilities in hopes of convincing them to use this form. If a facility prefers to use their own form, modifications will be made to the MTR as necessary to accommodate that facility.

The MTR database itself has undergone several structural changes over the course of year 01 in preparation for working with the MQIP facilities. The procedures for obtaining, validating and linking data and report generation have been modified to work in a more efficient manner thereby allowing the participation of a greater number of facilities in the registry. Current work with the CQI pilot facility is testing the new MTR procedures.

The certification component of the MQIP requires continuing education for mammography technologists as well as the radiologists. To meet this requirement, research has begun to determine the appropriate format and agenda for the technologist training. To this end, a focus group with mammography technologists located at the MQIP pilot facility was conducted. Technologists participating in this focus group described their vision of a useful training session that would include hands on training, information on breast pathology, training for special situations such as a handicapped patient, and MQSA requirements among other topics. The first technologist training event will occur in year 02 of the project.

Another training session that will be conducted as part of the MQIP is for state Certified Tumor Registrars (CTR's). The purpose of this training is to focus on their retrieval of data from medical records and to increase the quality of the data, specifically the recording of TNM staging and extent of disease information. It is anticipated that this training will improve the quality of the cancer data that is linked to the MTR and will thereby improve the content of the reports generated for the MQIP radiologists by providing more reliable and detailed size, stage and grade data. Preparation for this training event has been underway and it will happen in November, 1997.

CONCLUSIONS

Year 01 of the project has primarily been spent preparing for the implementation of the various components that make up the MQIP. Because the project is still in its initial phases, there are few conclusions to present.

Work that has been completed includes the design of the CQI as described in the paper in Appendix A. The design of the study incorporates measuring sensitivity and specificity rather than ROC curves. The statistical analysis that was involved with designing the CQI study demonstrates that sensitivity and specificity are much better measures of performance for mammography, especially using the universally accepted BIRADS coding scheme. This work is a significant contribution to the field of mammography quality and the ongoing discussion of the validity of ROC curves for measuring mammography and performance.

References

[1] Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. NEJM, 1994; 331(22):1493-1499.

[2] Moskowitz M. Guidelines for screening for breast cancer: is a revision in order? Radiological Clinics of North America 1992(Jan);30(1):221-233.

[3] Pisano ED, McLelland R. Implementation of breast cancer screening (review). Current Opinion in Radiology. Aug 1991; 3(4):579-87.

[4] Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S. Report of the international workshop on screening for breast cancer. National Cancer Institute, Bethesda, Maryland, February, 1993.

[5] Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. Can Med Assoc J 1992; 147:1459-1476

[6] Nystrom L, Rutgvist LE, Wall S, Lindren A, Lingvist M, Ryden S, Andersson I, Bjurstam N, Sagerberg G, Frisell J, Tabar L, Larsson LG. Breast cancer screening with mammography overview of Swedish randomized trials. Lancet 1993;341:973-978.

[7] Curpen BN, Sickles EA, Sollitto RA, Ominsky SH, Galvin HB, Frankel SD. The comparative value of mammographic screening for women 40-49 years old versus women 50-64 years old. AJR, May 1995;164(5):1099-103.

APPENDIX A

Design of a Study to Improve Accuracy in Reading Mammograms

Design of a Study to Improve Accuracy in Reading Mammograms

Margaret Sullivan Pepe,[1] *Nicole Urban,*[1] *Carolyn Rutter,*[2] *and Gary Longton*[1]
[1]DIVISION OF PUBLIC HEALTH SCIENCES, FRED HUTCHINSON CANCER RESEARCH CENTER, SEATTLE, WASHINGTON 98104,
AND [2]CENTER FOR HEALTH STUDIES, GROUP HEALTH COOPERATIVE, SEATTLE, WASHINGTON 98101

ABSTRACT. This paper is concerned with the design and analysis of mammography reading studies. In particular we consider studies aimed at evaluating interventions to improve the accuracy with which mammograms are read. A simple randomized design is suggested in which a relatively large group of readers read sets of mammograms before and after an intervention phase. We propose solutions to three difficult statistical issues which arise in the context of such studies: (i) the choice of primary outcome measure; (ii) the data analysis technique to be employed; and (iii) the methodology for calculating sample sizes for readers and images to be read.

First, we argue in favor of using sensitivity and specificity as the primary outcome measure rather than receiver operating characteristic (ROC) curves, although the latter are considered state of the art for many types of radiology reading studies. We argue that sensitivity and specificity are more clinically relevant and conceptually more straightforward than ROC curves. Second, we suggest a bivariate approach to data analysis for evaluating intervention effects on sensitivity and specificity. This accommodates the correlations inherent between these measures and allows for estimation of joint effects on them. Finally we propose a method for power calculations which utilizes computer simulation techniques. Simple formulas for sample size calculations are not available in part because variability in accuracy amongst readers and variation in difficultly amongst images introduce complexity into power calculations. The simulation method which we propose accommodates such complexity and is easy to implement.

The methodology was motivated by a study funded by the Department of Defense to evaluate the potential efficacy of an educational intervention. In the context of this study we illustrate the steps involved in power calculations and apply the data analytic techniques to the sort of data expected to result from this study. Though the proposed methods were motivated by this particular study, the statistical considerations are relevant more broadly in mammography and indeed in other types of radiologic imaging studies. Standards for the conduct of radiologic reading studies are not yet well developed, as they are for randomized clinical trials and for case-control studies. We hope that the discussion in this paper will add to the dialogue necessary for development of such standards.

KEY WORDS. ROC curves, sensitivity and specificity, computer simulation, diagnostic tests, screening.

1. INTRODUCTION

Mammography screening for breast cancer has been shown to be associated with decreased breast cancer mortality, at least in women over the age of 50 years [1]. Major efforts are currently underway to improve participation by women in screening programs [2]. Nevertheless, there is concern about the quality of mammography screening and there is general agreement that improvements in quality may lead to improvements in the performance of mammography as a screening modality. Quality might be improved for example by improving the imaging procedures. Alternatively, improvements in the accuracy with which mammographers interpret mammograms may improve the performance of screening mammography. Recent studies [3,4] have shown that there is considerable variability amongst radiologists in their interpretations of screening mammograms. Elmore et al [3] observed that sensitivities ranged from 74% to 96% and that specificities ranged from 35% to 89% among 10 radiologists reading 150 selected mammograms. Beam et al [4] using a much larger sample of 108 radiologists, each reading 79 mammograms, found sensitivities in the range of 47% to 100% and specificities in the range of 35% to 99%. These observations suggest that improvement in interpretation may be possible.

As part of a project called the Mammography Quality Improvement Project (MQIP) funded by the Department of Defense and aimed at improving the quality of mammography screening in rural communities, we are developing an educational program to improve the accuracy with which radiologists interpret mammograms. The educational intervention is comprised of a series of five sessions in which mammographers read films and are provided with immediate feedback on the accuracy of their interpretations. Feedback is provided using a laptop personal computer which is

mailed to the radiologist prior to his reading session. The computer program emphasizes the particular features of each mammogram which are relevant to determining the disease status of the woman screened. Eventually it may be possible to disseminate this sort of intervention over computer networks thus making it attractive in terms of easy accessibility and low cost.

In order to evaluate the impact of such an intervention on improvements in diagnostic accuracy it will eventually be necessary to perform a study of radiologists interpretations of screening mammograms in their actual practices. As a preliminary step to such a large-scale study, we will evaluate the intervention effects in a more controlled setting. Specifically, we will have a number of radiologists read a selected set of mammograms before and after the intervention and evaluate changes in accuracy. The mammograms included in this controlled study will be comprised of about 50% from women with disease, a proportion much larger than would be observed in practice but necessarily high in order to estimate sensitivity rates in a small-scale study. Mammograms will be selected to represent a reasonably broad range of interpretive difficulty.

The purpose of this paper is to elucidate some of the key statistical issues in the design of such a controlled reading study. Standards for the design of such studies are not well developed. This contrasts with therapeutic clinical trials and epidemiologic studies where the basic elements of study design are now fairly well standardized [5]. The question we propose to address in this reading study, namely evaluation of an intervention effect in a controlled setting, is a standard sort of question addressed in diagnostic imaging research. Hence the design issues which are dealt with here will have implications for future studies in mammography and in other diagnostic

test settings. These same issues also arise in reading studies designed to compare different imaging modalities. The key issues concern the choice of relevant primary outcome measures, appropriate data analysis strategies, and methodology for power calculations which incorporates variability amongst radiologists and amongst images. Broader issues in regards to study designs for evaluating imaging tests have been discussed in a more general sense in the literature [6,7].

In Section 2 we consider two sets of measures which can be used to define accuracy in reading mammograms; firstly, sensitivity and specificity and secondly, ROC curves. We argue in favor of the former, in part, because they are more clinically relevant and most easily understood but also because the latter can provide inappropriate conclusions concerning intervention benefits. In Section 3 we detail the basic elements of the statistical design of our study which could be considered a prototype for evaluating intervention effects in diagnostic radiology. An approach to joint analysis of sensitivity and specificity is outlined in Section 4. In Section 5 we describe methodology for power calculations which are appropriate for the proposed design and analysis. We propose the use of computer simulation methods for calculating power because they allow for complex designs and can easily incorporate variability amongst radiologists and images. Having described the steps involved in calculating power in Section 5, we then apply these procedures to the proposed MQIP study in Section 6, in order to illustrate the methods. Concluding remarks follow in Section 7.

2. MEASURES OF ACCURACY

2.1 Definitions

A radiologist reading a set of mammograms for a woman in our study will classify each

breast according to his or her suspicion of its showing malignancy. The ACR lexicon for rating a breast [8] which we will employ, defines a 5-point scale with category 1 indicating "normal, routine follow-up recommended", 2 indicating "benign, routine follow-up", 3 indicating "probably benign, early recall recommended", 4 indicating "suspicious for cancer, consider biopsy" and 5 indicating "highly suspicious for cancer, biopsy recommended". A common definition of a screen positive mammogram is one which receives a rating of 4 or greater. These are mammograms which are sufficiently suspicious for cancer that biopsy is recommended and hence they have an impact on clinical practice. Sometimes a rating of 3 or greater is considered positive. Because of the clinical implications of ratings 4 and 5, we will focus on the positivity criterion of category ≥ 4 here.

Given a definition for screen positivity, since there is a rating for each breast, one can calculate sensitivities and specificities with either 'woman' or 'breast' as the unit of analysis. The latter includes all non-diseased breasts, (including non-diseased breasts from women with cancer), as the denominator for specificity and all diseased breasts as the denominator for sensitivity. Breast level definitions are arguably inappropriate in clinical studies of mammography. It seems more clinically relevant to use woman as the unit of analysis. For this, one could use the maximum of the ratings for the left and right sides as the woman level rating for calculation of sensitivity and specificity. However, occasionally a woman with unilateral disease may not have it detected in the affected side but will have a positive mammogram on the unaffected side. In this case, using the maximum rating will inappropriately inflate the sensitivity. We define sensitivity instead as the proportion of women with disease who have it detected (a rating of ≥ 4) on the affected side. The specificity is

the proportion of women without disease who have a maximum rating of less than 4.

ROC analysis is a statistical technique used to describe accuracy of diagnostic tests when the test outcome is either ordinal or continuous as opposed to binary. The rating data generated in radiology reading studies is ordinal and ROC analysis is often considered optimal for the analysis of such studies as is evidenced, for example, in a recent issue of *Academic Radiology* [9]. An ROC curve is constructed by varying the criterion used for defining a positive mammogram from "rating ≥ 2" to "rating ≥ 5", plotting the associated sensitivity and 1-specificity values against each other, and finally fitting a curve to the points so that the curve is anchored at (0,0) and (1,1). Various algorithms exist for fitting a curve, the most notable being the Dorfman-Alf algorithm based on the binormal model [10] and the empirical nonparametric method which simply connects observed ROC points linearly. The area under the ROC curve is usually used to summarize accuracy. Again we suggest that woman rather than breast should be the unit of analysis in defining the ROC curve. That is, in calculating the sensitivity corresponding to the criterion "rating $\geq K$", it should be defined as the proportion of women with cancer who have a rating of $\geq K$ on an affected side.

2.2 ROC Analysis Versus Sensitivity and Specificity

ROC analysis was developed originally for diagnostic tests with results on some arbitrary scale. Its primary advantage is that it allows one to assess the inherent capacity of the test to distinguish between diseased and non-diseased subjects without linking the test to some particular threshold for defining screen positive [11,12]. This seems appropriate in radiology experiments when image ratings are arbitrary numbers with no specific clinical meaning attached to them. In that case, shifts in the distribu-

tions of ratings are of no consequence as long as they are equally shifted for diseased and non-diseased subjects. In mammography, however, mammogram ratings have very specific clinical meanings and consequent clinical implications. Uniform shifts in the frequencies with which rating categories are chosen can have major clinical implications.

Moreover, in contrast to the prototype setting for ROC analysis, shifts between certain diagnostic categories are of more importance than others. For example, as noted by Kopans [13], whether an image is rated in category 4 versus category 5 has no clinical impact. Similarly classifications in category 1 versus category 2 are clinically irrelevant. However, shifts between categories 4 or 5 and between 1 or 2 can have a big impact on the ROC analysis. To illustrate this consider the setting shown in Figure 1. The effect of intervention in this setting is to shift classifications of diseased observations from category 4 to category 5 and classification of non-diseased subjects from category 2 to category 1. Though these changes are of no clinical import, the ROC type analysis indicates a benefit for the intervention. Thus an ROC analysis can indicate a benefit of intervention even though a clinically relevant benefit does not exist.

Of even more concern is the fact that a clinically relevant benefit of intervention can occur even when the ROC curves pre- and post-intervention are the same. Consider the ROC curve depicted in Figure 2 for such a situation. The location on the ROC curve of the points associated with the criterion "rating $\geq$ category 4" indicate that sensitivity was significantly increased without decreasing specificity. This clinically relevant improvement in test accuracy does not manifest itself in an improvement in the ROC curves since the pre-and post-intervention curves are the

same. (Interestingly, classic binormal ROC curves do not fit the situation depicted in Figure 2 and a binormal ROC analysis in this setting may incorrectly indicate that the ROC curve post-intervention is improved over that pre-intervention).

The fact that ROC analysis can yield inappropriate conclusions regarding the clinically relevant effects of intervention argues against its use for the primary analysis of mammography reading study data. Another valid argument for not using an ROC analysis is that it is complicated and not easily understood by clinicians. Moreover, the so-called 'area under the curve' which summarizes the ROC curve in a single number has an interpretation which is not well known or easily understood. It can be interpreted as the probability that a radiologist will have a greater suspicion of cancer from a mammogram from a woman with disease than from a woman without [14]. This probability, however, seems to be of more theoretical than practical relevance.

We propose using the more clinically meaningful quantities of sensitivity and specificity for the primary data analysis and employing ROC analysis as a secondary descriptive device. Though ROC analysis may be statistically more powerful in some settings, any study should be designed so that it has adequate power to detect changes in the quantities which are of practical relevance. Hence, we suggest that power calculations for a mammography reading study should be based on the ability to detect changes in sensitivity and specificity rather than on the basis of detecting changes in ROC curves.

3. STUDY DESIGN

We now describe the basic elements of the design which we propose for studies evaluating intervention effects on reading accuracy in mammography. In this prototype

design, radiologists are randomly assigned to intervention and control groups, with the number in the former being denoted by R_T and the number in the latter denoted by R_C. Two image sets are constructed with M images in each set $S = 1, 2$. In set S, a number M_D^S are from women with disease and this number may differ between the two sets. Each reader reads one set of images before the intervention period and one set after. It is important that the sets before and after intervention be different since readers may remember, to some degree, images that they have previously read. Half of the readers chosen at random in each of the intervention and control groups read set 1 before intervention and set 2 after intervention. The other half read them in the opposite order: set 2 followed by set 1. This cross-over of film sets eliminates the possibility of systematic bias due to film sets. The design is balanced in the sense that set 1 is read equally often before and after the intervention phase in both the intervention and control groups, and similarly for set 2. Readers are told the approximate prevalence of diseased images, i.e., $(M_D^1 + M_D^2)/2M$ and that this varies between the two sets. The rationale for telling the readers the approximate prevalence is that it will become apparent in any case after reading the first set of images and that *apriori* knowledge of it should reduce the potential impact as much as possible on the observed improvement in accuracy. Readers will use the ACR lexicon to classify mammograms and for each reading it will be determined if it is screen positive or negative according to whether the rating is greater than 3 or not.

Images for inclusion in the study need to be selected so that average sensitivity and specificity at the baseline assessment are relatively low. That is, improvements in accuracy should be possible with the sets of images chosen. If, in the absence of intervention all images from women with disease were easily identified as such, the

observed sensitivities pre- and post-intervention would be close to 1 and a change in sensitivity would not be identifiable regardless of the actual effect of intervention. Thus at least some of the diseased images should be difficult but not impossible to identify as being from women with disease. Analogous considerations apply to specificity and the choice of non-diseased images included in the study.

4. DATA ANALYSIS

Having described the basic elements of the design and the choice of primary outcomes, we turn now to the strategy for data analysis. There are two components to the analysis. The first concerns a comparison of post- versus pre-intervention reading accuracy among the R_T readers in the intervention group. The second is the comparison of changes from pre- to post-intervention between the intervention and control groups. We first consider the former analysis, in part because it allows us to define notation most easily.

The purpose of this data analysis is to compare the overall sensitivity pre-intervention with that post-intervention and to compare the overall specificity pre-intervention with that post-intervention. If $\hat{S}_{r,\text{pre}}$ and $\hat{S}_{r,\text{post}}$ denote the observed pre- and post-intervention sensitivities for radiologist r, then the observed change in the overall sensitivity $\hat{\Delta}$(sensitivity) is the average change in sensitivities across radiologists in the intervention group:

$$\hat{\Delta}_T(\text{sensitivity}) = \frac{1}{R_T}\sum_{r=1}^{R_T}\left(\hat{S}_{r,\text{post}} - \hat{S}_{r,\text{pre}}\right).$$

Similarly the observed change in the overall specificity in the intervention group is

$$\hat{\Delta}_T(\text{specificity}) = \frac{1}{R_T}\sum_{r=1}^{R_T}\left(\hat{F}_{r,\text{post}} - \hat{F}_{r,\text{pre}}\right)$$

where $\hat{F}_{r,\text{pre}}$ and $\hat{F}_{r,\text{post}}$ denote the observed pre- and post-intervention specificities for radiologist r. Variance estimators for $\hat{\Delta}_T$(sensitivity) and $\hat{\Delta}_T$(specificity) are provided in the appendix. Although $\hat{\Delta}_T$(sensitivity) and $\hat{\Delta}_T$ (specificity) are sample means of changes in sensitivities and specificities, their variances are not given by the usual variance formulae for sample means. Indeed such sample variances would overestimate the variability. Rather the correct variance estimators rely on acknowledging that there are in essence two strata of radiologists in the design, which are defined by the ordering of the two image sets which are rated. The variances of $\hat{\Delta}_T$(sensitivity) and $\hat{\Delta}_T$(specificity) are averages of stratum-specific variances, as shown in the appendix.

Sensitivity and specificity are highly correlated parameters. Radiologists with high sensitivities tend to have low specificities. This will happen for example if they have a low threshold for classifying images as diseased. Similarly, changes in sensitivities and specificities induced by the intervention may be highly correlated. In particular, if the intervention simply changes the implicit threshold a radiologist has for classifying a mammogram as diseased then the sensitivity and specificity will both be changed but in opposite directions. Thus it is important to assess joint effects of intervention on sensitivity and specificity and to account for correlations between them in making inference. This can be accomplished by employing a bivariate analysis approach which is a special case of multivariate analysis, and for which there is a large statistical literature [15]. Using this approach to test the hypotheses that the true average sensitivity and specificity are unchanged by the intervention, H_0 : Δ(sensitivity) = Δ(specificity) = 0, a chi-squared test statistic is calculated. This statistic is a function of the observed average changes, $\hat{\Delta}$(sensitivity)

and $\hat{\Delta}$(specificity), their variances and also their correlation. An expression for the chi-squared statistic is provided in the appendix.

In addition to simply testing the hypothesis of no intervention effect, it will be important to provide a confidence region for the intervention effects on sensitivity and specificity based on the observed data. That is, a range of intervention effects, $\{\Delta(\text{sensitivity}), \Delta(\text{specificity})\}$, which are consistent with the observed data. Such a joint 95% confidence region is defined formally as the set of values (x, y) for which the hypothesis H_0 : $\{\Delta(\text{sensitivity}) = x, \Delta(\text{specificity}) = y\}$ is not rejected at the 5% significance level. This region is an elipse, centered at the observed intervention effect ($\hat{\Delta}$(sensitivity), $\hat{\Delta}$(specificity)). We refer the interested reader to the text [15] by Johnson and Wichern (1988, section 5.2) for technical details regarding its calculation. Code for calculating such regions has been written by Murdoch and Chow for the S-PLUS statistical software package and can be obtained from the S-archive on the Statlib computer site (http://lib.stat.cmu.edu). In a similar fashion a joint confidence region for the overall average sensitivity and specificity pre- or post-intervention can be calculated. It is calculated using the observed radiologist specific sensitivities and specificities pre- and post-intervention, and requires only calculation of the means, variances and correlations for these parameters. In order to illustrate these analyses, Figure 3 displays joint confidence regions based on a simulated data set. In our opinion these confidence regions provide a simple summary of the information contained in study data regarding intervention effects on reading accuracy. In the simulated data, the analyses show that sensitivity was increased by the intervention whereas there is no evidence of change in specificity.

So far we have considered the comparison of post- versus pre-intervention reading

accuracy within the intervention group. In order to attribute changes in accuracy to the intervention it will be necessary to compare the changes in the intervention group with those in the control group. Without the control group comparison observed changes might be attributed to other factors, such as the increased reading practice or increased awareness of reader fallibility induced by participation in the study. Thus, turning now to the comparison of intervention and control groups, the main hypothesis to be tested is that the changes in sensitivity and specificity in the intervention group are the same as those in the control group. Using a subscript T to denote the intervention group and subscript C to denote the control group, the null hypothesis is $H :_0 \ \Delta_C(\text{sensitivity}) = \Delta_T(\text{sensitivity}), \Delta_C(\text{specificity}) = \Delta_T(\text{specificity})$. A test statistic which has a chi-squared distribution with 2 degrees of freedom is described in the appendix for testing this hypothesis. Joint confidence regions for the differences in changes between the groups, namely $\Delta_T(\text{sensitivity}) - \Delta_C(\text{sensitivity})$ and $\Delta_T(\text{specificity}) - \Delta_C(\text{specificity})$, can be calculated using methods analogous to those described earlier for the pre-versus-post-intervention comparison.

5. METHODOLOGY FOR POWER CALCULATIONS

Power calculations for the reading study are somewhat complicated. They must accommodate the facts that readers vary in their accuracy parameters of sensitivity and specificity, that their sensitivities and specificities are likely negatively correlated, that images vary in difficulty and that a bivariate analysis approach will be employed. These factors together make analytic expressions for sample size intractable. We instead take a computer simulation approach to power calculations. The simulation approach to power calculation is a general and standard method and indeed software has been developed for certain types of applications [16]. The basic idea is to re-

peatedly simulate data as it is expected or hoped to arise in the course of the study, and determine how often the null hypothesis is rejected. By definition the statistical power of the study is the proportion of simulated studies in which the null hypothesis is rejected. One calculates the power in this fashion using various sample sizes until a sample size is found which provides adequate power. This indirect computer intensive approach to sample size calculation is easily accomplished with modern computers.

5.1 Models for Pre- and Post-intervention Accuracy

In order to simulate study data we need to define precisely the mechanisms giving rise to the data. We therefore need to make assumptions about the reading accuracies before and after intervention. For this purpose we suppose that before intervention a reader correctly assesses a woman with tumor as being diseased with probability $P^D_{r,i}$. The probability $P^D_{r,i}$ depends on the image denoted by i and on the reader, denoted by r. The probabilities $P^D_{r,i}$ will presumably be higher if the tumor is clearly visible in image i than if it is not. The probabilities will also be higher if the radiologist is conservative and is inclined to recommend biopsy for borderline cases. We let S^D be the sensitivity of the average radiologist to the average film from a woman with tumor. The variability amongst films in terms of the difficulty which readers have in assessing them, is captured by specifying a distribution for the sensitivities that the average reader has in assessing the films. Here we assume that the average reader's sensitivity to films varies uniformly in an interval $(S^D - a^D, S^D + a^D)$ across different films. Thus for the average radiologist, easier films are read with sensitivity closer to $S^D + a^D$ and more difficult films are read with sensitivity closer to $S^D - a^D$. In a similar fashion, on the average film from a diseased woman, the sensitivity of different readers is assumed to vary uniformly in an interval $(S^D - b^D, S^D + b^D)$

across radiologists. Thus radiologists with high sensitivity to the average film will have sensitivity closer to $S^D + b^D$. In the appendix we detail a logistic model with random effects (also called a mixed model) for the probabilities $P^D_{r,i}$ which give rise to inter-image and inter-reader variability as postulated here. It is assumed that on the logistic scale there are no interactions between reader and image specific effects on the sensitivity.

Observe that for the purposes of simulating data, by specifying S^D and a^D we can now generate a random image effect by choosing a random number in $(S^D \pm a^D)$ which corresponds to the sensitivity an average radiologist has for detecting it. Similarly, having a specified S^D and b^D we are in a position to generate a random reader effect by choosing a random number in $(S^D - b^D, S^D + b^D)$ which corresponds to his sensitivity to the average film. The logistic model displayed in the appendix then yields the probability $P_{i,r}$ which that reader has of correctly assessing that image as diseased.

Analogous considerations apply to the determination of randomly generated specificities which vary across radiologists and across images from women without disease. Values for parameters $S^{\bar{D}}$, $b^{\bar{D}}$ and $a^{\bar{D}}$ need to be specified in order to define the data generating process. Here, $S^{\bar{D}}$ is the probability that the average radiologist will correctly assess the average non-diseased image as such, radiologists vary uniformly in $(S^{\bar{D}} - b^{\bar{D}}, S^{\bar{D}} + b^{\bar{D}})$ in their specificities to the average non-diseased film, and images from women without disease vary uniformly in $(S^{\bar{D}} - a^{\bar{D}}, S^{\bar{D}} + a^{\bar{D}})$ in the probabilities of the average reader correctly classifying them. The sensitivities and specificities from single radiologists should be correlated. In the appendix we describe how negative correlation between sensitivities and specificities within radiologists can

be built into the data simulation mechanism.

In summary, for each study radiologist we simulate his/her sensitivity and specificity to the average diseased and non-diseased films respectively, by randomly sampling correlated numbers from $(S^D - b^D, S^D + b^D)$ and $(S^{\bar{D}} - b^{\bar{D}}, S^{\bar{D}} + b^{\bar{D}})$, respectively. For each study film we determine the sensitivity or specificity that an average radiologist has for it by randomly sampling a number from $S^D - a^D, S^D + a^D)$ or $(S^{\bar{D}} - a^{\bar{D}}, S^{\bar{D}} + a^{\bar{D}})$. Finally, for each combination of film i and radiologist r, we can calculate $P^D_{i,r}$ or $P^{\bar{D}}_{i,r}$, which is the probability that the radiologist will assess that image correctly.

The $P^D_{i,r}$ and $P^{\bar{D}}_{i,r}$ pertain to probabilities before intervention in the treatment and control groups. One also needs to specify treatment effects in order that corresponding probabilities after intervention can be calculated. We postulate that after intervention the quantities S^D and $S^{\bar{D}}$ are changed to new values but that the variations amongst readers and amongst images remain the same. In the appendix we define in a mathematically precise way a logistic model which incorporates such intervention effects.

5.2 Simulated Study Data Generation

Having specified statistical models for pre- and post-intervention rating probabilities which incorporate variation amongst radiologists and amongst images, we now turn to the simulation of study data in accordance with the study design which we proposed in section 3. The first step is to generate images and image sets. This entails generating M diseased images (i.e., M image-specific parameters, one for each image), generating M non-diseased images, and finally from the $2M$ films choosing M at random without

replacement to form film set 1. The remaining M films constitute film set 2. The next step is to generate R_T intervention readers and R_C control readers and assign them film sets. That is, for each of $R_T + R_C$ readers we generate pairs of pre- and post-intervention sensitivities and specificities to average diseased and non-diseased films according to the models described in section 5.1. Of the total $R_T + R_C$ readers, R_T are assigned at random to the intervention group and the remaining R_C to the control group. Finally film set orderings are assigned to the readers with half of the intervention readers selected at random being assigned set 1 first and the other half assigned set 2 first. Similarly, $R_C/2$ control readers are assigned set 1 followed by set 2 and the other $R_C/2$ readers are assigned film sets in the opposite order.

The final step in generating data for a simulated study is to actually generate the readings for each reader and image combination. That is, for each reader and for each of the M films in his/her pre-intervention set, a binary random variable is generated which is his/her assessment of whether or not that image shows disease using the probability $P^{D}_{r,i,\text{pre}}$ if the image is diseased and $1 - P^{\bar{D}}_{r,i,\text{pre}}$ if the image is not diseased. Similarly, for each of the M films in his/her post-intervention set a similar binary random variable is generated using $P^{D}_{r,i,\text{post}}$ or $1 - P^{\bar{D}}_{r,i,\text{post}}$ noting that the pre- and post-probabilities differ by different amounts for intervention-versus-control radiologists.

Having generated the simulated study data the test statistics of interest can now be calculated. Data are simulated (first the probabilities, then the ratings) and results calculated under the same assumptions and study design many times, with 1000 or 5000 simulated datasets being typical numbers used for power calculations. The proportion of simulated studies in which the null hypothesis is rejected is the

calculated study power for that design and under those assumptions.

6. POWER CALCULATIONS: RESULTS FOR THE MQIP STUDY

To fix ideas, we now illustrate the computer simulation method for power calculations in the MQIP study. This illustration also identifies some sources of data to guide assumptions for power calculations.

We need to choose assumed parameters for the baseline sensitivities and specificities, for the variations amongst radiologists and amongst images and for intervention effects of interest. We assume that the median sensitivity pre-intervention, S^D, in our study will be in the range of .70 to .80. This accords with previous studies which found median sensitivities of .70 and .80 [3,4]. Median pre-intervention specificity will also be assumed to lie in the range of .70 to .80. Beam et al [4] found a median specificity of 0.94 for mammograms from women with normal mammograms and a median specificity of 0.60 for mammograms from women with benign disease. Elmore et al [3] found a median specificity of 0.94. In contrast to these studies we will inform the radiologists of the average prevalence which is higher than that expected in a practical screening setting. Because of this and the fact that the films in our study will be somewhat difficult, we anticipate an initial specificity lower than observed in those studies. The variation amongst radiologists in sensitivities and specificities will be assumed such that $b^D = 0.20$ and $b^{\bar{D}} = 0.20$, which is in agreement with the range of approximately 40% in sensitivities (and specificities) amongst radiologists observed in Beam's study. We could find no data on inter-image variability to suggest appropriate values for a^D and $a^{\bar{D}}$. We assume that they are of the same order of magnitude as the inter-rater variability parameters, $a^D = a^{\bar{D}} = .20$. In regards to intervention effects of interest, we consider that changes of 10 percentage points

in either sensitivity or specificity are of interest. However, we calculated power for a variety of intervention effects.

Practical considerations concerning time and cost dictate the range of sample sizes which are feasible and therefore, for which power calculations are performed. We anticipate that no more than approximately 80 radiologists are available for the reading study in the rural communities in which our mammography quality improvement study is being conducted. To maximize power, equal numbers of radiologists are assigned to control and intervention groups. Therefore the number of radiologists per group to be considered for power calculation purposes will be in the range of 20 to 40. Experience suggests that readers can comfortably read no more than 45 films per session. We therefore calculated power for experiments in which the number of films per set, M, was either 30 or 45.

Estimates of power based on computer simulations are shown in Table 1. Though results are shown only for intervention effects on sensitivity with no effect on specificity, because of the symmetry inherent in the design, the same power calculations hold for a 10% change in specificity with no change in the sensitivity. Observe that the power is far larger for the within intervention group assessment of change than for the between group comparison of change. This is to be expected since the variability involved in comparing two random changes is greater than the variability involved in comparing a single change with the null hypothesis of no change. We also observe from Table 1 that the power is less when the baseline sensitivity is 0.70 than when it is 0.80. This is due to the relatively larger binomial variance for the lower baseline rate. In order to be conservative we focus on this lower rate. Interestingly the baseline specificity had little impact on the power to detect an intervention effect on the

sensitivity.

The target power for our study design is 90% which allows a 10% chance of an inconclusive result when the intervention increases sensitivity from .70 to .80. For the within intervention group comparison this cannot be achieved with 20 readers but can be achieved with 30 readers if 45 images are included in each image set. The between group comparison, however, has a power of only 66% in this case. Even with use of our maximum resources, i.e., 40 readers per group and 45 images per reading set, the power is only 80%. This allows for a 20% chance of an inconclusive result even when there is a clinically important intervention effect on diagnostic accuracy.

For the MQIP study we chose to focus the study on the within group comparison. The power calculations were an important contribution to this decision but other considerations also played a role. Radiologists would have little motivation to participate in the control arm whereas they would receive CME credit for participation in the intervention arm. The possibility that those in the control arm would learn from the baseline assessment was also a concern and thus we were concerned that it might not even be feasible to construct a true control group. Finally, it was felt that if we found a definite positive change in the intervention group, then this would provide sufficient motivation to proceed with more comprehensive controlled studies in the future. Thus we chose to study only the intervention effects in the intervention group and to use sample sizes of 30 radiologists each reading sets of mammograms from 45 women before and after intervention.

The simulation program allowed us the flexibility to explore the performance of this study design in a variety of settings other than that assumed for the primary sample size calculation. First we calculated the probability of rejecting the null

hypothesis for settings where there was no intervention effect. Recall that inference for the test statistic is based on a chi-squared statistic and is theoretically valid with large samples. However, this study entails relatively small samples. We used the simulations to check the adequacy of the large sample theory in our study. To do this we generated data under the null hypothesis. The rejection probability was approximately .06 in the settings we studied, indicating that the true significance level of the test is slightly higher than the target of .05 but adequate for our purposes.

We next explored the power of this study design and sample sizes to detect an array of intervention effects. Results are shown in Table 2. Although the study has adequate power to detect a change in sensitivity (or specificity) of 0.10 even when the pre-intervention sensitivity is as low as 0.60, it has little chance of detecting a smaller change of 0.05. On the other hand, if small changes of the order of .05 occur in both the average sensitivity and in the average specificity there is a good chance that the simultaneous effects will be detected.

7. DISCUSSION

Diagnostic imaging technology is already a basic component of medical care and continues to develop at a rapid pace. It is clearly important to assess the accuracy with which readers can diagnose disease using such technologies, to evaluate the effects of training strategies and to compare methods. Implications for public health can be enormous. Unfortunately, statistical methodology for evaluating and comparing imaging methods has not received much attention by biostatisticians and epidemiologists involved in public health research. Rather the literature is concentrated in radiology research journals, has generally focused on small scale studies involving only a few readers and has ignored clinical implications associated with different di-

agnostic categories. We believe that it is time to bring the discussion about study design and analysis for evaluating imaging technology to the broader community of epidemiologists and statisticians involved in public health. This is particularly important as interest increases in the accuracies and costs of these imaging methods. By presenting our thoughts on the design and analysis of a study to evaluate an educational intervention on the interpretation of mammograms, we hope to stimulate such discussion.

The choice of primary outcome measure is the most basic element of any study design. We chose to consider the sensitivity and specificity as the basis for evaluating intervention effects. This conflicts with initial statistical reviewers of our study design who were of the opinion that ROC analysis was the only appropriate and indeed the state-of-the-art basis for evaluating an intervention effect. We now argue that in mammography where specific clinical actions are associated with diagnostic rating categories, sensitivity and specificity provide a more clinically relevant and conceptually straightforward basis for comparison than does ROC analysis. Moreover this approach allows us to evaluate effects on false positive as well as true positive rates. In contrast ROC analysis does not quantify the false positive rates directly but in a sense only uses it to standardize the true positive rate. We do not dismiss ROC analysis entirely but rather we regard the analysis of the specific rating categories of secondary importance and focus the design on sensitivity and specificity. Thus the MQIP study was designed to ensure adequate power to detect changes in the most clinically relevant quantities.

We also needed to decide upon the analysis techniques for making statistical inference about sensitivity and specificity. We propose to simultaneously estimate

sensitivity and specificity using multivariate methods. Sensitivity and specificity as we have defined them are average sensitivities and average specificities of radiologists in our study. They can also be interpreted as marginal or population average quantities, in the sense of being the probability that a diseased (or non-diseased) image will be correctly interpreted as such in the study. The distinction between the population average and average radiologist-specific interpretations has to do with whether one considers the accuracy parameters to be based on data pooled across radiologists (population average) or to be based on calculation of the accuracy parameter for each radiologist and then averaging the results. In our study these quantities coincide because all radiologists expect to read the same numbers of films. In studies where this is not the case, the distinction should be considered and a decision should be made regarding which of the two entities is most relevant.

The approach we propose for statistical inference is relatively straightforward, being based on methods for inference about sample means. Confidence intervals are based on the variance-covariance matrix of the estimated (sensitivity, specificity) parameters or their changes amongst radiologists. Possible non-normality of the average estimates may be an issue in our study, though for the settings considered in the power calculation this did not appear to be the case. An alternative approach to inference which might be more robust would follow the marginal regression modelling approach described by Leisenring, Pepe and Longton [17]. One could formulate logistic regression models for the population average sensitivity and 1-specificity as

$$\text{logit } \{\text{Prob[screen positive} \mid \text{image diseased]}\} = \gamma_0 + \gamma_1 b$$

$$\text{logit } \{\text{Prob[screen positive} \mid \text{image non-diseased]}\} = \eta_0 + \eta_1 b$$

where the logit function is $\text{logit}\{x\} = ln\{x/(1-x)\}$ and b is 0 if the image was read

before the intervention and 1 if it was read after the intervention. The changes in the true and false positive rates are now quantified in the odds ratio parameters γ_1 and η_1, respectively, and joint confidence intervals can be calculated. By adding an interaction term between b and I, where I is an indicator of the radiologist being in the control or intervention groups:

$$\begin{aligned} \text{logit } \{\text{Prob[screen positive} \mid \text{image diseased}]\} &= \gamma_0 + \gamma_1 b + \gamma_2 bI \\ \text{logit } \{\text{Prob[screen positive} \mid \text{image non-diseased}]\} &= \eta_0 + \eta_1 b + \eta_2 bI \end{aligned}$$

a comparison of the changes in the intervention and control groups can be made by testing if the parameters γ_2 or η_2 are 0. Though this logistic regression modelling approach may provide more robust confidence intervals, we felt that the simpler approach described earlier was adequate for power calculations.

The prototype reading study we have described concerns evaluating the effect of an intervention on the change in accuracy parameters. We note, however, that most of our discussion is also relevant to the comparison of accuracies associated with different imaging modalities. Suppose for example, that there are two sets of women (denoted by set 1 and set 2) from which images have been made using two modalities. A natural study design to compare the modalities would entail readers assigned to read one set of films produced with one modality and the other set of films produced with the other modality. Using the notation $1(A)$ to denote set 1 produced with modality A and similarly for the other combination, readers read either $\{1(A)$ and $2(B)\}$ or $\{2(A)$ and $1(B)\}$. Considering that the ordering may also influence accuracy parameters, this yields four groups of readings, $\{1(A), 2(B)\}$, $\{2(B), 1(A)\}$, $\{2(A), 1(B)\}$ and $\{1(B), 2(A)\}$. A balanced cross-over design would assign radiologists randomly to these four reading assignments. The difference in the sensitivity and specificity

between modality A and B can be calculated by simply pooling all relevant readings for modality A and similarly for modality 2. Inference for the difference follows in the same fashion as that described for the change induced by intervention in the intervention group of our study but that now there are 4 rather than 2 strata of radiologists defined by the image reading set assignments.

Power calculations for reading studies are not straightforward due in part to correlations induced by images and readers. That is, for each image there are multiple readings. Moreover, each reader provides multiple readings and radiologist specific sensitivities and specificities are correlated. We propose valid and simple analyses for dealing with these factors but power calculations required a computer simulation approach. We found the process of developing the computer simulation study to be a useful exercise. It compels one to think through the processes generating study data. It also allows one to experiment with the assumptions and design easily. For example, we considered designs which included a larger number of film sets to be read in the study and found that the study power was decreased slightly due to the extra variation introduced. Computer simulations also allow one to check how test statistics perform under the null hypothesis with sample sizes proposed in the study. Hence one can check if inference based on large sample theory is valid in the setting where it is to be applied. We suggest that simulation studies are a useful approach to power calculations in any setting, though given the complexities in radiology reading studies, the case for the technique in this setting is particularly strong.

ACKNOWLEDGMENTS: We appreciate the support of grants CA63145 and GM54438 awarded by NIH, and grant ~~MM950024~~ DAMD17-96-1-6288 awarded by the Department of Defense. We thank Molly Edmonds for her excellent technical help in preparing the

manuscript.

References

1. Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC. Report on a workshop of the UICC Project on Evaluation of Screening for Cancer. **International Journal of Cancer** 1990; 46: 761-769.

2. Rakowski W, Andersen MR, Stoddard AM, Urban N, Rimer BK, Lane DS, Fox SA, Costanza ME for the NCI Breast Cancer Screening Consortium. A confirmatory analysis of the pros and cons of mammography. **Health Psychology,** in press.

3. Elmore JG, Wells CK, Lee CH, Howard DH and Feinstein AR. Variability in radiologist's interpretation of mammograms. **New England Journal of Medicine** 1994; 3331: 1493-1499.

4. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists: Findings from a national sample. **Archives of Internal Medicine** 1996; 156: 209-213.

5. Begg CB. Advances in statistical methodology for diagnostic medicine in the 1980's. **Statistics in Medicine** 1991; 10: 1887-1895.

6. Begg CB and McNeil BJ. Assessment of radiologic tests: control of bias and other design considerations. **Radiology** 1988; 167: 565-569.

7. Gatsonis C and McNeil BJ. Collaborative evaluations of diagnostic tests: experience of the Radiology Diagnostic Oncology Group. **Radiology** 1990; 175: 571-575.

8. Kopans DB, D'Orsi CJ, Adler DD, Bassett LW, Brenner RJ, Dodd GD, Feig SA, Lopiano MA, McLelland RM, Moskowitz M and Sickles EA. **American College of Radiology Breast Imaging Reporting and Data System, 1993.**

9. Advances in Statistical Methods for Diagnostic Radiology: A Symposium. **Academic Radiology** 1995; 2:S1.

10. Dorfman DD, Alf E. Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals: rating method data. **Journal of Mathematical Psychology** 1969; 6: 487-496.

11. Swets JA and Pickett RM. **Evaluation of Diagnostic Systems: Methods from Signal Detection Theory.** 1982; Academic Press, New York.

12. Metz CE. ROC methodology in radiologic imaging. **Investigative Radiology** 1986; 21: 720-733.

13. Kopans DB. The accuracy of mammographic interpretations. **New England Journal of Medicine** 1994; 331: 1521-1522.

14. Hanley JA and McNeil BJ. The meaning and use of the area under operating characteristics curve. **Radiology** 1982; 143: 29-36.

15. Johnson RA and Wichern DW. **Applied Multivariate Statistical Analysis.** 1988; Prentice Hall, New Jersey.

16. EGRET Siz module. Statistics and Epidemiology Research Corporation, Seattle, Washington, USA.

17. Leisenring W, Pepe MS and Longton GL. A marginal regression modelling framework for evaluating medical diagnostic tests. **Statistics in Medicine**, (in press).

APPENDIX

1. Variance Estimators for Change in Overall Sensitivity and Specificity.

The change in the overall sensitivity defined in Section 4 can be written formally mathematically as

$$\hat{\Delta}_T(\text{sensitivity}) = \frac{1}{R_T}\left\{\sum_{r:(\text{order}=1,2)}(\hat{S}_{r,\text{post}} - \hat{S}_{r,\text{pre}}) + \sum_{r:(\text{order}=2,1)}(\hat{S}_{r,\text{post}} - \hat{S}_{r,\text{pre}})\right\}$$

where $\hat{S}_{r,\text{pre}}$ is the observed sensitivity for radiologist r with his pre-intervention film set and $\hat{S}_{r,\text{post}}$ is the corresponding quantity post-intervention. Observe that the order of film sets essentially defines two strata in this setting and the notation (order $= 1,2$) (or (order $= 2,1$)) used to denote the stratum in the summation indicates that it includes only radiologists assigned sets in the order set 1 first and set 2 second (or set 2 first and set 1 second). The variance of $\hat{\Delta}_T$(sensitivity) can be estimated using the variance of a stratified sample mean $\hat{V} = (\hat{V}_{(1,2)} + \hat{V}_{(2,1)})/R_T$, where $\hat{V}_{(1,2)}$ is the sample variance of the quantities $(\hat{S}_{r,\text{pre}} - \hat{S}_{r,\text{post}})$ in the stratum (order=1,2), and $\hat{V}_2$ is the analogous quantity in the other stratum. The ratio $\hat{\Delta}_T(\text{sensitivity})/\sqrt{\hat{V}}$ can be compared with a standard normal distribution to test for a change in the sensitivity which is statistically significantly different from 0.

2. Chi-squared Test Statistics for Bivariate Analyses.

To simultaneously test the null hypotheses that both the sensitivity and specificity are unchanged in the intervention group, $H_0:\ \Delta_T(\text{sensitivity}) = 0 = \Delta_T(\text{Specificity})$, the following test statistic can be used

$$\left[\hat{\Delta}_T(\text{sensitivity})\ \hat{\Delta}_T(\text{specificity})\right]\ \hat{\sum}_T^{-1}\begin{bmatrix}\hat{\Delta}_T(\text{sensitivity})\\ \hat{\Delta}_T(\text{specificity})\end{bmatrix}$$

where the square bracket notation is used to denote vectors and $\hat{\Sigma}_T^{-1}$ is the inverse of a square matrix $\hat{\Sigma}_T$. This matrix $\hat{\Sigma}_T$ is a variance-covariance matrix for the two-dimensional statistic [$\hat{\Delta}_T$(sensitivity) $\hat{\Delta}_T$(specificity)], and is the analogue of the variance $\hat{V}$ defined above in relation to the one-dimensional quantity $\hat{\Delta}_T$(sensitivity). Formally we write

$$\hat{\Sigma}_T = \left\{ \hat{\Sigma}_T^{(1,2)} + \hat{\Sigma}_T^{(2,1)} \right\} \Big/ (R_T - 1)$$

where $\hat{\Sigma}_T^{(1,2)}$ is the sample variance-covariance matrix for the quantities $\left\{ \hat{S}_{r,\text{post}} - \hat{S}_{r,\text{pre}}\ \hat{F}_{r,\text{post}} - \hat{F}_{r,\text{pre}} \right\}$ in the stratum (order = 1,2), and $\hat{\Sigma}_T^{2,1}$ is the analogous quantity calculated for the other stratum. The test statistic is compared with a standard chi-squared distribution with 2 degrees of freedom in order to test the null hypothesis concerning changes in sensitivities and specificities.

Consider now the component of the data analysis concerning the comparison of changes between intervention and control groups. Using a subscript C to denote the control group in analogy with our use of the subscript T to denote the intervention group, we define the statistics $\hat{\Delta}_C$(sensitivity), $\hat{\Delta}_C$(specificity) and $\hat{\Sigma}_C$. The estimated differences between the groups in changes of sensitivities and specificities can be written as $\hat{\Delta}_T$(sensitivity) $-$ $\hat{\Delta}_C$(sensitivity) and $\hat{\Delta}_T$(specificity) $-$ $\hat{\Delta}_C$(specificity), respectively. The hypothesis that the changes are the same for intervention and control groups can be tested by comparing the statistic

$$\left[\hat{\Delta}_T(\text{sens}) - \hat{\Delta}_C(\text{sens})\ \hat{\Delta}_T(\text{spec}) - \hat{\Delta}_C(\text{spec}) \right] \left[\Sigma_T + \Sigma_C \right]^{-1} \begin{bmatrix} \hat{\Delta}_T(\text{sens}) - \hat{\Delta}_C(\text{sens}) \\ \hat{\Delta}_T(\text{spec}) - \hat{\Delta}_C(\text{spec}) \end{bmatrix}$$

with the quantiles of a chi-squared distribution with 2 degrees of freedom, where we use the abbreviations 'sens' and 'spec' to denote 'sensitivity' and 'specificity' in the above expressions.

3. Mixed Models for Reading Accuracies.

Section 5 outlines a statistical model for sensitivity and specificity parameters which vary with reader and image. Here we present a more formal and precise definition of this model. For radiologist r on diseased film i, we write the chance of correctly identifying it as diseased pre-intervention using a logistic model as

$$P^D_{r,i} = \exp\{\mu^D + \gamma^D_i + \beta^D_r\} / \left(1 + \exp\{\mu^D + \gamma^D_i + \beta^D_r\}\right)$$

where γ^D_i and β^D_r are random variables specific to this film and radiologist, respectively. For the average radiologist $\beta^D_r = 0$, and for the average film $\gamma^D_i = 0$. Thus for the average radiologist on the average film the sensitivity is $S^D = \exp\{\mu^D\}/(1 + \exp\{\mu^D\})$. The films vary in difficulty in the sense that the average radiologist has a lower sensitivity on some films and a higher sensitivity on others. Mathematically this translates into allowing γ^D_i to vary. We choose it as a random variable so that the average radiologist's sensitivity to different films varies uniformly in an interval $(S^D - a^D, S^D + a^D)$. Technically this is achieved by letting $\gamma^D_i = \ln\{U^D_i/(1 - U^D_i)\} - \mu^D$ where U^D_i is a random variable with a uniform distribution in $(S^D - a^D, S^D + a^D)$. The radiologists also vary amongst themselves in their sensitivities to the same film and this inter-rater variation translates into allowing β^D_r to vary. We simulated data so that on the average diseased film (i.e., $\gamma^D_i = 0$) the sensitivities of radiologists varied uniformly in $(S^D - b^D, S^D + b^D)$. Again, technically we let $\beta^D_r = \ln\{U^D_r/(1 - U^D_r)\} - \mu^D$ where U^D_r is a random variable with a uniform distribution on the interval $(S^D - b^D, S^D + b^D)$.

Turning now to specificities, we write the specificity for radiologist r on non-

diseased film j pre-intervention as

$$P^{\bar{D}}_{r,j} = \exp\{\mu^{\bar{D}} + \gamma^{\bar{D}}_j + \beta^{\bar{D}}_r\} \Big/ (1 + \exp\{\mu^{\bar{D}} + \gamma^{\bar{D}}_j + \beta^{\bar{D}}_r\})$$

where in analogy with the above notation for diseased films, the average radiologist on the average film has specificity $S^{\bar{D}} = \exp\{\mu^{\bar{D}}\}/(1 + \exp\{\mu^{\bar{D}})$ and parameters $a^{\bar{D}}$ and $b^{\bar{D}}$ indicate variation in the specificity with film and radiologist. As argued in section 5, data should be generated so that the β^{D}_r and $\beta^{\bar{D}}_r$ are negatively correlated. We incorporated this into the simulation by first generating the sensitivity radiologist-specific random effect parameter, β^{D}_r, (i.e., his/her sensitivity to the average film) which is based on the random variable U^{D}_r, and then letting the corresponding random variable for the specificity random effect be defined as

$$U^{\bar{D}}_r = \left\{ \left(S^{\bar{D}} - (U^{D}_r - S^{D}) \frac{b^{\bar{D}}}{b^{D}} \right) \right\}.$$

Thus if the radiologist's sensitivity is $x \times b^{D}$ above the average radiologist's sensitivity to the average film, S^{D}, his/her specificity will be $x \times b^{\bar{D}}$ below the average specificity to the average film.

Our model postulates that after intervention the quantities $S^{\bar{D}}$ and S^{D} are changed to new values but that the radiologist and image-specific parameters remain unchanged. Thus, suppose that after intervention the sensitivity of the average radiologist to the average film is $\exp(\mu^{D} + \alpha^{D}\}/(1 + \exp\{\mu^{D} + \alpha^{D}\})$. Then the chances that radiologist r will correctly classify film i pre- and post-intervention are

$$P^{D}_{r,i,\text{pre}} = \exp\{\mu^{D} + \gamma^{D}_i + \beta^{D}_r\}/(1 + \exp\{\mu^{D} + \gamma^{D}_i + \beta^{D}_r\})$$

and

$$P^{D}_{r,i,\text{post}} = \exp\{\mu^{D} + \alpha^{D} + \gamma^{D}_i + \beta^{D}_r\}/(1 + \exp\{\mu^{D} + \alpha^{D} + \gamma^{D}_i + \beta^{D}_r\}),$$

respectively. Similarly the postulated change in $S^{\bar{D}}$ specifies a parameter $\alpha^{\bar{D}}$ (analogous to α^{D}) which facilitates calculation of post-intervention specificities. Having chosen values for the various parameters $(\mu^{D}, \alpha^{D}, a^{D}, b^{D})$ and $(\mu^{\bar{D}}, \alpha^{\bar{D}}, a^{\bar{D}}, b^{\bar{D}})$, this completes the first step of the simulation power calculation method, namely specification of accuracy parameter distributions pre-intervention and intervention effects.

Table 1. Study power to detect a 10% increase in sensitivity with no accompanying effect on specificity pre- versus post-intervention within the intervention group and power to detect a 10% increase in sensitivity in the intervention group versus no change in the sensitivity in the control group when specificites in both groups remain unchanged after the Intervention. All tests are two sided and are tested at a significance level of .05.

Readers Per Group (R_T)	Films Per Set (M)	Pre-intervention Sensitivity	Pre-intervention Specificity	Power	
				Within Intervention Group	Comparison with Control Group
20	30	.70	.70	.70	.38
20	30	.70	.80	.66	.34
20	30	.80	.70	.79	.45
20	30	.80	.80	.77	.44
20	45	.70	.70	.81	.48
20	45	.70	.80	.82	.53
20	45	.80	.70	.91	.61
20	45	.80	.80	.92	.64
30	30	.70	.70	.81	.48
30	30	.70	.80	.83	.52
30	30	.80	.70	.93	.60
30	30	.80	.80	.91	.61
30	45	.70	.70	.94	.66
30	45	.70	.80	.95	.66
30	45	.80	.70	.99	.80
30	45	.80	.80	.99	.79
40	30	.70	.70	.92	.61
40	30	.70	.80	.94	.60
40	30	.80	.70	.97	.73
40	30	.80	.80	.98	.75
40	45	.70	.70	.98	.79
40	45	.70	.80	.99	.80
40	45	.80	.70	.99	.88
40	45	.80	.80	.99	.89

Table 2. Study Power to detect various configurations of changes in the intervention group using a study design with 30 readers and 45 films per set. The pre-intervention specificity is assumed to be .70 in all cases. The intervention induced change in sensitivity is denoted Δ_T(sens) and in specificity is denoted Δ_T(spec).

Pre-Intervention Sensitivity	Δ_T(sens)	Δ_T(spec)	Power
.60	+0.10	0.00	0.90
.70	+0.10	0.00	0.95
.80	+0.10	0.00	0.98
.60	+0.05	0.00	0.35
.70	+0.05	0.00	0.39
.80	+0.05	0.00	0.50
.60	+0.05	+0.05	0.66
.70	+0.05	+0.05	0.68
.80	+0.05	+0.05	0.71

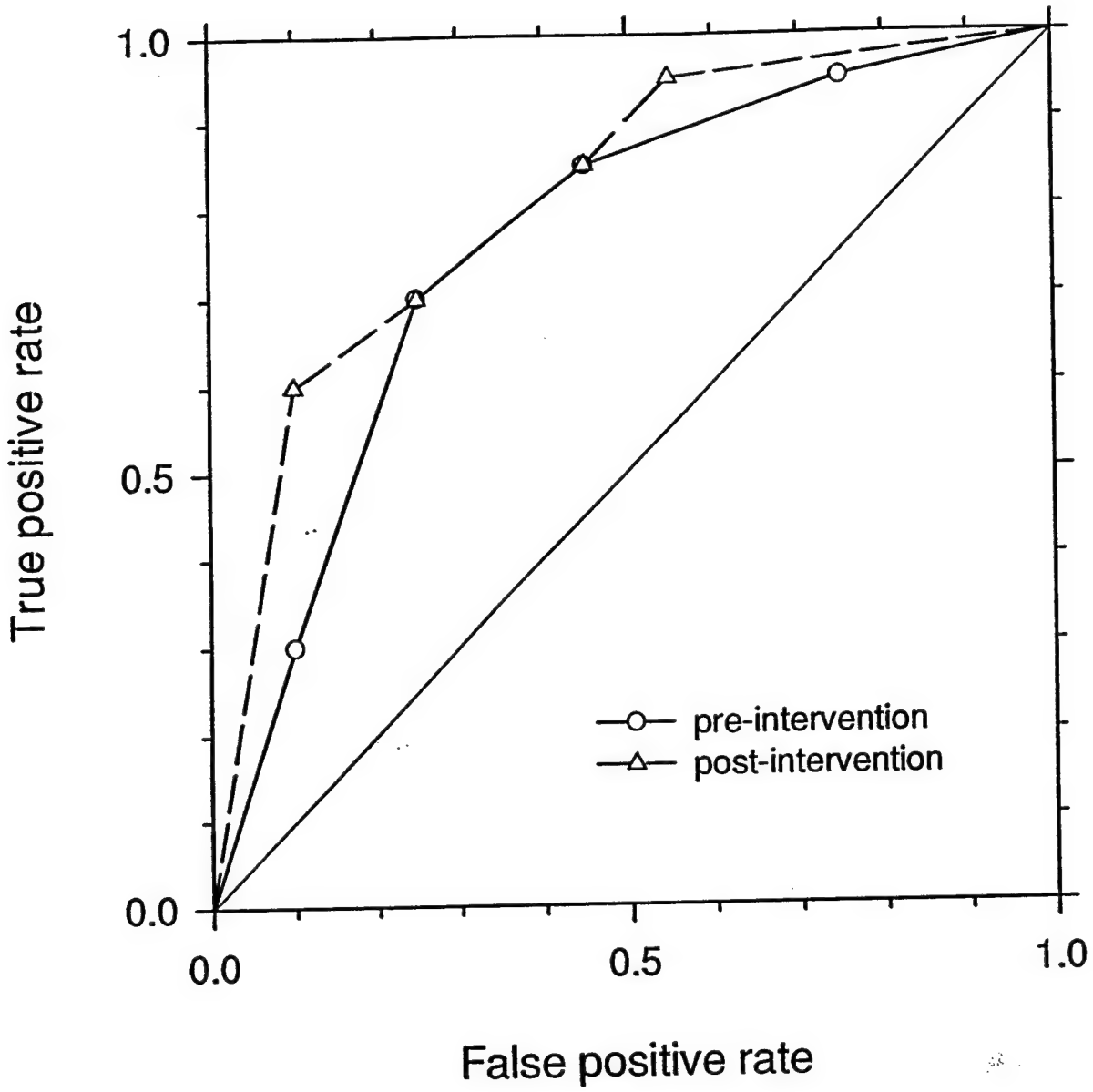


Figure 1: An hypothetical setting where the sensitivity and specificity associated with the clinically relevant criteria are unchanged but the empirical ROC curves indicate a benefit of intervention. The (false positive, true positive) points associated with categories 5, 4, 3, and 2 are (.10, .30), (.25, .70), (.45, .85), and (.75, .95) respectively pre-intervention and (.10, .60), (.25, 70), (.45, .85), and (.55, .95) respectively post-intervention.

9/23/96

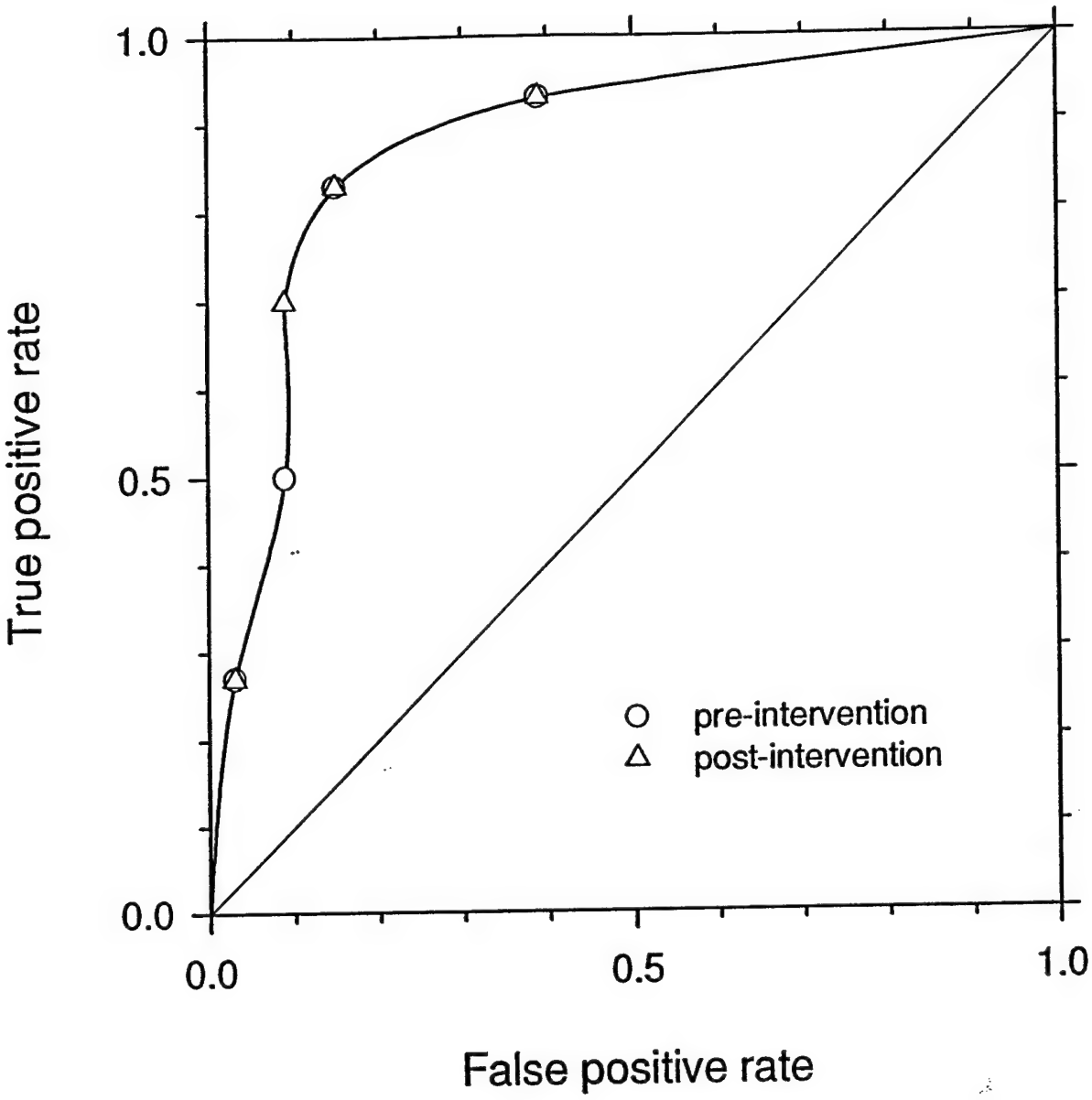
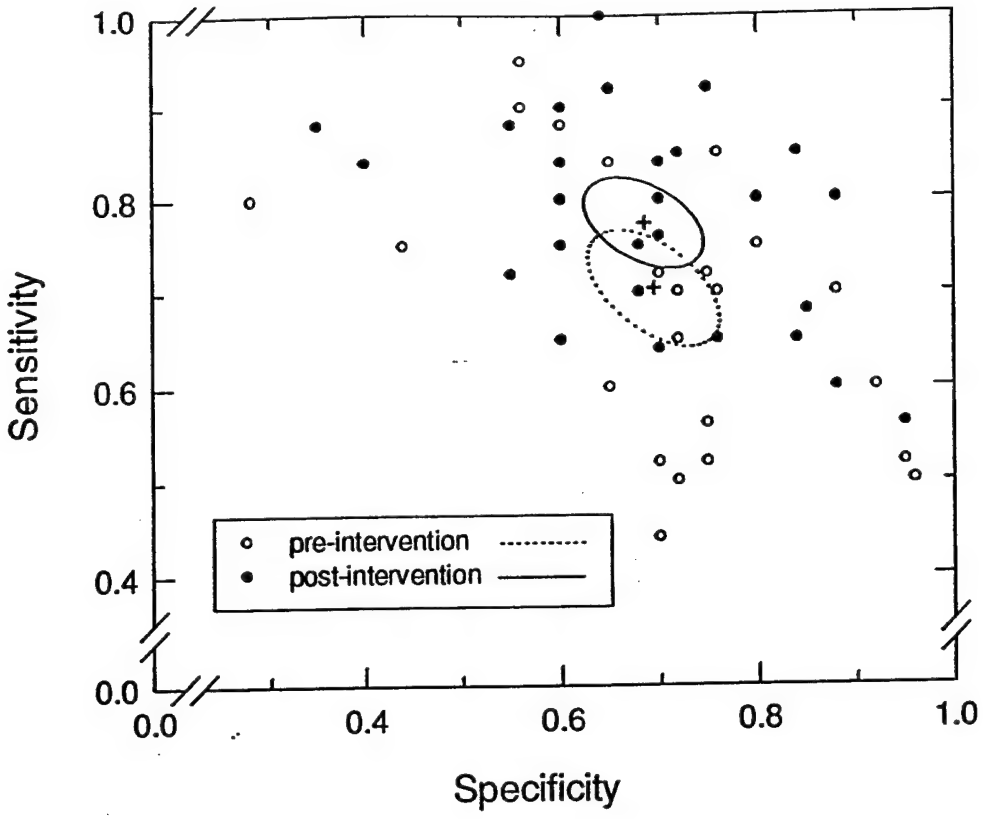


Figure 2: An hypothetical setting where ROC curve is unchanged by the intervention but there is a clinically relevant benefit. The sensitivity associated with the clinically relevant criterion is improved from .50 to .70 while the associated false positive rate remains unchanged at 0.09. The (false positive, true positive) points associated with categories 5, 4, 3, and 2 are (.03, .27), (.09, .50), (.15, .83), and (.39, .93) pre-intervention and (.03, .27), (.09, .70), (.15, .83), and (.39, .93) post-intervention. These points before intervention are labelled with circles and after intervention are labelled with triangles.

5/2/97



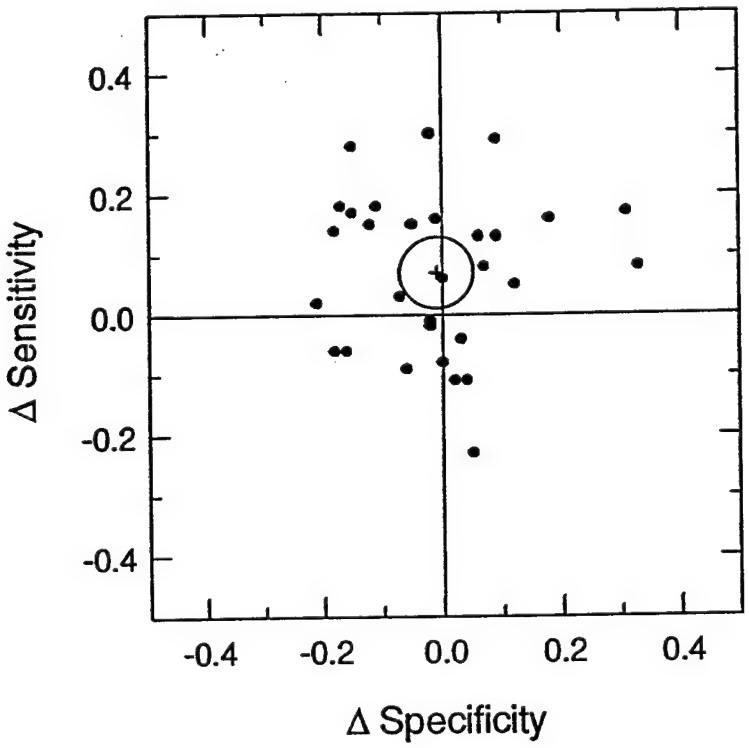


Figure 3: Joint confidence regions for sensitivity and specificity both pre and post intervention along with a joint confidence region for the changes in these parameters. Data used in this illustration were generated using computer simulation methods described in sections 5 and 6. Points correspond to observed data for individual radiologists.

APPENDIX B

Recruitment Procedures and Materials for the CQI Study

I Facility Recruitment

A. Overview

A minimum of sixteen rural mammography facilities or facility groups that when combined will have at least 60 radiologists reading mammograms for them, will be recruited to participate in the MQIP. The focus of the recruitment effort will be on smaller, more rural facilities or facility groups that do not have dedicated mammographers, e.g., radiologists who read only mammograms. A facility group is defined as a combination of facilities that

- have a business association; or
- may share radiologists.

Facility-associated radiologists may read for a second, unrelated facility outside of the facility group, however, the second clinic may not be part of another MTR/MQIP selected facility group.

After selection of the sixteen (or more) facilities targeted for the MQIP intervention, each facility will first be recruited to participation in the Mammography Tumor Registry (MTR), a state-wide effort to link facility mammograms with the state cancer registry. Each facility will be required to sign a Data Release/Confidentiality Agreement that allows the MTR to link mammograms performed at the facility with information in the Washington State Cancer Registry.

All radiologists reading for the sixteen targeted facilities will be notified after the facility signs the agreement. See Section II, Radiologist Recruitment for more details.

B. Recruitment Procedures

All targeted facilities will be recruited to the MTR in the same manner.

1. The first contact, a letter addressed to the facility/clinic manager, will consist of the following:
    - an introduction to the Mammography Tumor Registry (Appendix – Letter 1);
    - an MTR brochure outlining the benefits of participation in the registry (Appendix – promotional 1);
    - a return card (Appendix – Form 1) on which the facility indicates the method – paper or computer – used to collect and store patient information, health history, and examination information.

2. Approximately ten days after the initial mailing to the facility, a field coordinator will telephone to confirm the information on the returned card or to obtain the information if the card was not returned. In addition to confirming the information on the returned card, calls to facilities will also determine the following:
    - the correct person to receive future MTR/MQIP mailings (sample scripts, Appendix –Forms 2 and 3);
    - additional mammography facility sites associated with the one being contacted;

- for multiple facilities under one management, information about whether their database is shared or if individual databases are maintained at each site;
- a correct business office address and telephone number.

3. After the MTR receives the data collection methods' card (or after the facility is contacted by phone and the information is obtained), the appropriate information packet will be assembled and mailed. The information, presented in a folder, includes the following:
    - a cover letter to the facility manager or designated facility personnel that previews the MQIP and the letter that will go to the facility's reading radiologists offering them the opportunity to be part of the MQIP and the randomized CQI study. (Appendix – Letter 2a, 2b);
    - an MTR brochure (Appendix – Promotional 1);
    - two copies of the Data Release/Confidentiality Agreement – one to be kept for the facility files (Appendix – Form 4);
    - a facility staff directory and a radiologist directory for completion, correction and/or updating (sample, Appendix – Forms 5 and 6);
    - questions and answers about the MTR, either computer or paper data collection specific (Appendix – Forms 7a and 7b)
    - facilities that collect and store data on computer will receive a sample of the validation report (sample, Appendix – Form 8)
    - for facilities that collect and store information on paper only, a draft copy of a new data collection form that they are urged to use because it requests information needed for linking (Appendix – Form 9); computerized facilities receive a copy to confirmed the desired data to be collected;
    - further, paper-based facilities have the opportunity to personalize the paper data collection form with their facility name by returning a card indicating the heading they would like included on their forms (Appendix – Form 10);
    - A stamped return envelope.

4. If there is no response from the facility, e.g., no return of the documents, within ten working days, a field coordinator will telephone to answer questions and check on the status of the paperwork. At this time, the field coordinator will make certain the packet arrived and offer to send a new one if necessary. Any questions about MTR that cannot be answered by the field coordinators will be directed to the appropriate MTR staff.

5. After receipt of the signed Data Confidentiality Form, the facility is considered part of the MTR. The appropriate MTR data staff will begin working with the facility to extract data from their files so that the linking process with the Washington State Cancer Registry data can begin.

II Radiologist Recruitment

A. Overview

All radiologists reading for the targeted participating facilities or facility groups will be the focus of recruitment to the MQIP. After the facility returns the corrected radiologist and staff directories, the information in the MTR FACTS system is updated and the radiologists' recruitment begins.

At that time the radiologists will be informed about the MQIP and urged to sign the participation agreement that allows potential randomization to the CQI program. Later all radiologists will be invited to submit interesting cases for inclusion in the CQI. Finally, participating radiologists will be offered the opportunity to sign the Radiologist's Data Release form. A radiologist's signature on the form allows him to receive regular, patient-identified reports on the mammograms he reads. When retrospective data becomes available, the radiologist will receive information that identifies patients for whom he read a mammogram who were later diagnosed with breast cancer.

B. Recruitment Procedures

All radiologists reading for the MTR targeted facilities will be recruited in the same manner.

1. An initial letter to be sent to each targeted radiologist will include the following information:
    - an introduction to the MQIP with emphasis on their associated facility's interest in participation (Appendix – Letter 3);
    - the MQIP Consent to Participate Agreement which details the project and the randomization process (Appendix – Form 11);
    - A stamped, return envelope.

2. Whether or not the signed consent is returned, each targeted radiologist will receive a second mailing ten days after the first. This mailing will contain the following:
    - a cover letter which reminds the radiologist to sign and return the Participation Agreement if it hasn't been returned yet;
    - a request for interesting mammography films to be considered for possible inclusion in the CQI software program (Appendix – Letter 4);
    - four copies of the mammography Case Selection forms (Appendix – Form 12);
    - a stamped return envelope.

3. If there is no response, e.g., the signed Participation Agreement is not received ten working days after the second mailing, the field coordinators and/or the project co-investigator will telephone the radiologist to
    - offer to send the form again,
    - offer to answer questions about the MQIP or direct questions to the appropriate staff person, and
    - urge the radiologist to submit interesting mammography films for possible inclusion in the CQI.

The radiologist will receive reminder calls, or a second mailing, until the signed agreement is returned or the radiologist declines further participation.

4. After all targeted radiologists have either returned the signed participation agreement (or declined participation), randomization to the CQI will take place. Radiologists (N = >30) at one-half of the participating facilities will be randomized to take part in the Continuous Quality Improvement (CQI) program. After randomization, all radiologists will be notified about their status in the CQI program and their opportunity to receive the Radiologist's Quality Improvement Report.

5. The notification letter for radiologists randomized to the CQI will include:
    - a thank you for participation, information about the randomization, an introduction to the MTR emphasizing the benefits for radiologists, and a preview of the first CQI scheduling phone call (Appendix – Letter 5a);
    - an MTR brochure (Appendix – Promotional 1);
    - a Radiologist Data Release form (Appendix – Form 13);
    - an MQIP timeline - *in development* (Appendix – Form 14);
    - a stamped addressed envelope.

6. For radiologists not randomized to the CQI, the letter will include:
    - a thank you for participation and an introduction to the MTR emphasizing the benefits for radiologists (Appendix – Letter 5b);
    - a Radiologist's Data Release form (Appendix – Form 13);
    - a stamped addressed envelope.

The radiologist's signature on the data release form allows the radiologist to begin receiving the Radiologist's Quality Improvement report that identifies patients for whom he/she read a mammogram and who were later diagnosed with breast cancer. These reports will be sent when there is sufficient data in the data base to link the patients through the state cancer registry. While it is desirable that the radiologist agree to receive the confidential report, a radiologist does not have to sign the data release form in order to participate in the CQI.

III The Continuous Quality Improvement Sessions

A. Overview

The CQI sessions are designed to determine the following:

1. determine inter-rater variability in film interpretation before and after the implementation of the program which tests the radiologists on a set of mammograms selected specifically for their teaching value;
2. determine post-CQI intra-rater variability in film interpretation;
3. determine if digitized mammograms can be interpreted with the same accuracy as high-quality copies of mammograms;
4. determine if the accuracy of film interpretation depends on covariates such as the woman's age and breast density.

B. Procedures

The CQI sessions are intended to occur over a 12 - 14 month period with every effort made to have a consistent and approximately equal amount of time (8 –12 weeks) between sessions.

1. To schedule the first session, the field coordinators will phone each physician's office and schedule either with the physicians individually or with the office manager. Every effort will be made to schedule all physicians in a practice during the same field coordinator visit. Office staff will be informed of the necessary requirements for the sessions including
   - A separate room;
   - space to set up the light box;
   - space for the radiologist to sit to use the laptop computer;
   - two electrical outlets.

The CQI sessions will involve the radiologist reading 45 (sessions 1 and 4) or 30 (sessions 2 and 3) high quality copies of mammography films. He/she will enter an assessment of each case on a digitized image of the film on a laptop computer. A field coordinator will facilitate each session by

- introducing the session;
- giving basic instructions;
- positioning the films on the light box;
- giving instructions for the CQI computer program; and
- ensuring that the sessions follow project protocol.

2. Before the first session, the physician will receive:
   - a reminder postcard indicating the day/date/time of the session to be mailed about one week in advance (Appendix – Form 15);
   - a reminder phone call two days before the session.

3. If a physician has not signed the radiologist data release form, he will be offered the opportunity to do so at the first CQI session. Field coordinators will carry the forms

with them and offer to answer, or direct to any MTR staff person, questions about the radiologist report.

4. At each session beginning with the first, the following session will be scheduled. The field coordinator will remind the physician about the timeline, which is approximately 8 – 10 weeks between sessions, and schedule the next session on her calendar. Follow-up will include
    - A reminder postcard mailed two weeks prior to each of sessions 2 – 5 which allows the physician time to reschedule if necessary;
    - a phone call will be made a day or two before the session as a courtesy reminder.

5. When courtesy is extended to field coordinators by the radiologists' staff or facility personnel, individual handwritten thank you notes will be sent. All correspondence will be entered into the FACTS system.

After all five CQI sessions are completed, the radiologist will receive certification notice of the 10 CME credits earned.

IV Appendix

A. Letters
B. Forms
C. Promotional Materials

Letter 1
Initial Facility Contact

September 18, 1997

«cfacnm»
Attn: «ccontact»
«caddr1»
«caddr2»
«ccity», «cstate» «czip»

Dear «ccontact»:

In a survey of mammography facilities that was conducted last fall, your facility expressed interest in the Washington State Mammography Tumor Registry (MTR). The MTR provides follow-up mammography information to facilities and offers researchers the opportunity to gain insight into issues related to breast cancer detection. In the enclosed brochure that summarizes the benefits of the MTR, you will note that a key benefit of participation is a report containing information that helps satisfy the audit recommendations of the Mammography Quality Improvement Act (MQSA).

Please complete and return the enclosed card indicating how your facility collects and stores patient and examination information. We will then send you an information packet and details about participation in the MTR.

As an MTR participant, your facility is also eligible for the Mammography Quality Improvement Project (MQIP). The MTR and the Mammography Quality Improvement Project are being conducted by investigators at the Fred Hutchinson Cancer Research Center. The MQIP is aimed at predominantly rural areas of Washington State where mammography use is being studied. In our ongoing effort to reduce breast cancer mortality, the Hutchinson Center was awarded a grant to examine methods for improving the quality of mammography.

The MQIP will offer a Continuing Medical Education conference for facility technologists at no cost to the facility or the technologist. Each radiologist who reads mammograms for your facility will also be contacted about the MQIP. For radiologists, a key component of the MQIP is participation in a Continuous Quality Improvement (CQI) program. This CQI program is designed to strengthen film-reading ability by providing immediate feedback on the radiologist's interpretation of a set of mammograms specifically selected for their teaching value. One-half of the participating radiologists will be randomly selected to take part in the CQI film-reading study. Initially, the CQI program will be available only in the context of this research study. Radiologists who are selected for the film reading study will receive Continuing Medical Education (CME) credits for their participation.

We look forward to your collaboration with us on the Mammography Tumor Registry and the Mammography Quality Improvement Project. If you have any further questions, please contact the project office at (800) 224-2331 or (206) 667-6560.

Sincerely,

Nicole Urban ScD
Principal Investigator

Enclosure: Brochure
Card
Stamped return envelope

00058

September 18, 1997

«cfacnm»
Attn: «ccontact»
«caddr1»
«caddr2»
«ccity», «cstate» «czip»

Dear «ccontact»:

We invite your facility to participate in the Washington State Mammography Tumor Registry (MTR). The MTR provides follow-up mammography information to facilities and offers researchers the opportunity to gain insight into issues related to breast cancer detection. In the enclosed brochure that summarizes the benefits of the MTR, you will note that a key benefit of participation is a report containing information that helps satisfy the audit recommendations of the Mammography Quality Standards Act (MQSA).

Please complete and return the enclosed card to indicate how your facility collects and stores patient and examination information. We will then send you an information packet and details about participation in the MTR.

As an MTR participant, your facility is also eligible for the Mammography Quality Improvement Project (MQIP). The MTR and the Mammography Quality Improvement Project are being conducted by investigators at the Fred Hutchinson Cancer Research Center. The MQIP is aimed at predominantly rural areas of Washington State where mammography use is being studied. In our ongoing effort to reduce breast cancer mortality, the Hutchinson Center was awarded a grant to examine methods for improving the quality of mammography.

The MQIP will offer a Continuing Medical Education conference for facility technologists at no cost to the facility or the technologist. Each radiologist who reads mammograms for your facility will also be contacted about the MQIP. For radiologists, a key component of the MQIP is participation in a Continuous Quality Improvement (CQI) program. This CQI program is designed to strengthen film-reading ability by providing immediate feedback on the radiologist's interpretation of a set of mammograms specifically selected for their teaching value. One-half of the participating radiologists will be randomly selected to take part in the CQI film-reading study. Initially, the CQI program will be available only in the context of this research study. Radiologists who are selected for the film-reading study will receive Continuing Medical Education (CME) credits for their participation.

We look forward to your collaboration with us on the Mammography Tumor Registry and the Mammography Quality Improvement Project. If you have any further questions, please contact the project office at (800) 224-2331 or (206) 667-6560.

Sincerely,

Nicole Urban, ScD
Principal Investigator

Enclosure: Brochure
Card
Stamped return envelope

00059

Letter 2a
Packet Cover Letter
Computer-based Facility

September 18, 1997

«cfacnm»
Attn: «ccontact»
«caddr1»
«caddr2»
«ccity», «cstate» «czip»

Dear «ccontact»:

Thank you for your interest in the Mammography Tumor Registry (MTR). The enclosed material includes a complete description of the MTR and the initial paperwork required for enrollment. After reading the project description, we ask you to do the following:

1. Sign and date both copies of the Data Release/Confidentiality Agreements. Keep a copy for your files and return one to the MTR. Our receipt of the signed agreement confirms your participation.

2. Review and complete or correct the Radiologist Directory and return the form to us. We need current information because each radiologist who reads mammograms for your facility will receive a letter of introduction to the Mammography Quality Improvement Project (MQIP). The introductory letter includes the Participation Agreement for the radiologist's signature.

3. Review and complete or correct the Facility Staff Directory so that our contacts are with the appropriate personnel. To help you determine which staff person should receive and review the data validation reports, we include a sample report.

4. In order to link mammography results with cancer incidence data, standardized patient and exam information are required. We include a copy of the patient history and medical record form used by MTR facilities that do not have a computerized system for data entry. To help us work with your facility, return the following forms:
    a. a copy of your patient information and medical history form with the data you enter highlighted and
    b. a copy of the exam form with all data that you currently enter highlighted.

If you have any questions after reading the information in this packet, please call 1(800) 224-2331 or (206) 667-6560 to speak to someone on our staff.

We are pleased to have you working with us on this valuable research endeavor.

Sincerely,

Nicole Urban, ScD
Principal Investigator

Letter 2b
Packet Cover Letter
Paper-Based Facility

September 18, 1997

«cfacnm»
Attn: «ccontact»
«caddr1»
«caddr2»
«ccity», «cstate» «czip»

Dear «ccontact»:

Thank you for your interest in the Mammography Tumor Registry (MTR). The enclosed material includes a complete description of the MTR and the initial paperwork required for enrollment. After reading the project description, we ask you to do the following:

1. Sign and date both copies of the Data Release/Confidentiality Agreements. Keep a copy for your files and return one to the MTR. Our receipt of the signed agreement confirms your participation.

2. Review and complete or correct the Radiologist Directory and return the form to us. We need current information because each radiologist who reads mammograms for your facility will receive letter of introduction to the Mammography Quality Improvement Project (MQIP). The introductory letter includes the Participation Agreement for the radiologist's signature.

3. Review and make necessary changes to the Facility Staff Directory so that our contacts are with the appropriate personnel. Return the corrected form.

4. To facilitate the collection of uniform data from all state mammography facilities, the MTR developed a standardized patient history and medical record form in accordance with the April 1996 Proposed Rules for the MQSA, Section 900.12. A copy of the form is enclosed. The MTR will print these personalized forms for your facility. Return the card attached to the form, indicating the information you would like included in the top margin. For example, you may request that the facility name or you may request white space for a patient identification label. The forms are provided in a carbon-less format; the facility keeps the original and returns a copy to the MTR on a regular basis.

If you have any questions after reading the information in the packet, please call 1(800) 224-2331 or (206) 667-6560 to speak to someone on our staff.

We are pleased to have you working with us on this valuable research endeavor.

Sincerely,

Nicole Urban, ScD
Principal Investigator

Letter 3
Radiologist's Initial Letter
Introduction to MQIP

September 18, 1997

«cpersfnm» «cperslnm», «ctitle»
«caddr1»
«caddr2»
«ccity», «cstate» «czip»

Dear Dr. «cperslnm»:

A facility for which you regularly read mammograms, «cgrpnm», has expressed interest in participating in the Mammography Quality Improvement Project (MQIP), a research study conducted by investigators at the Fred Hutchinson Cancer Research Center.

The Mammography Quality Improvement Project is aimed at predominantly rural areas of Washington State where mammography use is being studied. In our ongoing effort to reduce breast cancer mortality, the Hutchinson Center was awarded a grant to examine methods for improving the quality of mammography screening.

The attached participation agreement explains all aspects of the MQIP. One component of the MQIP is a research project that tests the efficacy of a Continuous Quality Improvement (CQI) film-reading program for radiologists. One-half of the participating radiologists will be randomly assigned to the CQI intervention group. In brief, the intervention consists of five two-hour film-reading sessions conducted over a 12 - 14 month period. In the first four CQI sessions, the radiologist reads a series of mammograms (cases) and enters an assessment of each case on a digitized computer image. At the fifth session, the radiologist reads and assesses digitized mammograms only. This last session will determine the feasibility of disseminating the CQI program via the Internet.

Films are read using standard assessment codes. On mammograms coded as possible cancer (3, 4 or 5), the radiologist indicates the location of the abnormality on the digitized image. Feedback is offered at the conclusion of each session and the radiologist is given the opportunity to review individual films. Ten hours of continuing medical education (CME) credits are awarded after the completion of the five sessions. All CME paperwork will be handled by our office.

Please read the attached participation agreement carefully, initial each page, complete and sign the final page, and return the agreement in the enclosed envelope. You will receive a copy for your files. After randomization occurs, you will be notified about your status. If you are randomized to the CQI intervention group, you will be contacted to schedule the first session in early 1998.

If you have any questions about the MQIP, please call the project office at (800) 224-2331 or (206) 667-6560. A staff person will be happy to help you.

Thank you for the prompt return of the enclosed paperwork.

Sincerely,

Nicole Urban, ScD
Principal Investigator

00062

Enclosure: MQIP Consent to Participate
Stamped Return Envelope

Letter 4
Radiologist's Reminder
Film Solicitation

September 18, 1997

«cpersfnm» «cperslnm», «ctitle»
«caddr1»
«caddr2»
«ccity», «cstate» «czip»

Dear Dr. «cperslnm»:

Recently you received a letter inviting you to participate in the Mammography Quality Improvement Project (MQIP). If you have not returned the participation agreement, please take a minute or two to sign and return it. We will mail you a signed and dated copy of the agreement.

You will recall that one aspect of the MQIP is the Continuous Quality Improvement (CQI) film-reading program. The radiologists who are randomly assigned to the CQI program will read 180 mammography films at five film reading sessions over a 12 – 14 month period.

The films used in the CQI sessions are being solicited from radiologists around the state. We hope you will submit 4-6 interesting cases for the pool of mammography films to be reviewed for study inclusion. We are looking for an equal number of benign and malignant assessments with about 25% from women under the age of 50.

Consider the following criteria when you review cases for submission:

- The woman was asymptomatic at the time of the mammogram.
- The interpretation was challenging.
- Malignant assessments have biopsy correlation.
- Benign assessment is determined either by 1) a woman's cancer free status two years after the mammogram; 2) additional evaluation resulting in a benign finding; or 3) a negative biopsy result.

Please complete one of the enclosed case selection forms for each case you submit and return them by ***insert date***. Do not send original mammograms. Our staff will contact the clinic storing the films and make the necessary arrangements to obtain them. All information you submit remains strictly confidential. Before the review and selection process, patient, clinic, and physician identifiers are masked on the films. Films chosen for study inclusion are masked during the copying and digitization procedures.

All films will be returned to the clinic within eight weeks. If there is an emergency, immediate return can be arranged by calling the project office at (800) 224-2331 or (206) 667-6560. If you have any questions about case submission, you can reach a project staff person at the above numbers.

Thank you for taking time to consider films for submission.

Sincerely,

Nicole Urban, ScD
Principal Investigator

Enclosure: Case Selection Forms
Stamped return envelope

00063

Letter 5a
Notification of Randomization Status
Receive CQI
Introduction to MTR

September 18, 1997

«cpersfnm» «cperslnm», «ctitle»
«caddr1»
«caddr2»
«ccity», «cstate» «czip»

Dear Dr. «cperslnm»:

Thank you for participating in the Mammography Quality Improvement Project (MQIP). You have been randomized to the Continuous Quality Improvement (CQI) film-reading intervention group. To remind you, the CQI sessions are about two hours in length and will be conducted on-site by a project field coordinator over a 12-14 month period. Please refer to the enclosed timeline for the intervention schedule. A field coordinator will contact you soon to schedule the first session.

You may also request information from the Mammography Tumor Registry (MTR). The benefits of the MTR are outlined in the enclosed brochure. For the radiologist, the MTR can provide a Radiologist Quality Improvement Report. These reports identify women for whom you read a mammogram who were later diagnosed with breast cancer. Facilities and radiologists are identified on the confidential reports by MTR assigned code only; no facility or radiologists' names appear. Because of the sensitive nature of the material, the reports come to you in an envelope stamped confidential.

This individual radiologist's report can be produced only after we receive the signed Radiologist Data Release form. Please read and sign the enclosed form and return it in the envelope provided. After data is received from the facility, undergoes a complete validation process, and is linked with the Washington State Cancer Registry, you will begin receiving the confidential Radiologist's Quality Improvement Report.

For mammography facilities, the MTR can provide a summary administrative report that helps satisfy the MQSA outcomes and audit recommendation. This facility level report contains no patient or physician identifiers.

If you have any questions about the CQI program or about the individual radiologist's report, please call (800) 224-2331or (206) 667-6560 to speak with someone on our staff.

Thank you. We will telephone soon to schedule the first CQI session.

Sincerely,

Nicole Urban, ScD
Principal Investigator

Enclosures: Radiologist's Data Release Form
MQIP Timeline
MTR Brochure
Stamped Return Envelope

00064

Letter 5b
Notification of no Randomization
Introduction to MTR

September 18, 1997

«cpersfnm» «cperslnm», «ctitle»
«caddr1»
«caddr2»
«ccity», «cstate» «czip»

Dear Dr. «cperslnm»:

Thank you for participating in the Mammography Quality Improvement Project (MQIP). You were not selected to participate in the randomized Continuous Quality Improvement (CQI) film-reading intervention group. We will notify you, however, when the CQI program is available on the Internet. The technologists at «cgrpnm» will be contacted about their opportunity to attend a Continuing Medical Education (CME) session at no cost to the facility or the technologist.

You may also request information from the Mammography Tumor Registry (MTR). The benefits of the MTR are outlined in the enclosed brochure. For the radiologist, the MTR can provide a Radiologist Quality Improvement Report. These reports identify women for whom you read a mammogram who were later diagnosed with breast cancer. Facilities and radiologists are identified on the confidential reports by MTR assigned code only; no facility or radiologists' names appear. Because of the sensitive nature of the material, the reports come to you in an envelope stamped confidential.

This individual radiologist's report can be produced only after we receive the signed Radiologist Data Release form. Please read and sign the enclosed form and return it in the envelope provided. After data is received from the facility, undergoes a complete validation process, and is linked with the Washington State Cancer Registry, you will begin receiving the confidential Radiologist's Quality Improvement Report.

For mammography facilities, the MTR can provide a summary administrative report that helps satisfy the MQSA outcomes and audit recommendation. This facility level report contains no patient or physician identifiers.

If you have any questions about the CQI program or about the individual radiologist's report, please call (800)224-2331or (206)667-6560 to speak with someone on our staff.

Thank you.

Sincerely,

Nicole Urban, ScD
Principal Investigator

Enclosures: Radiologist's Data Release Form
MTR Brochure
Stamped Return Envelope

00065

Form 2
Phone Contact Script
- Data Collection Method
Card Returned

Mammography Tumor Registry
Phone Contact A

Date: ___/___/___

Caller: _____

Facility Name _____

Address _____

Phone Number _____

Contact Person _____

Data Collection Card Returned? _____ yes *(after introduction, go to **section 1**)*
_____ no *(after introduction, go to **section 2**)*

Contact Notes:

_____
_____
_____
_____
_____
_____

INTRODUCTION

Hello. My name is [Marcia Gaul/Susie Wilson]. I'm calling from the Fred Hutchinson Cancer Research Center. Is the facility manager available to speak to me at this time?

1. *Yes:* Thank you.

2. *No: Arrange to leave a message or call back at a better time.*

Contact Name: _____
Title: _____
Phone Number: _____
Today's date: _____
Date to call back: _____

[redacted]

SECTION 1: ***Facilities that returned the data collection card***

Hello. My name is [Marcia Gaul/Susie Wilson] and I work at the FHCRC. I'm calling about the Mammography Tumor Registry, a research project we are conducting here at the Center. This is a follow-up to the letter we mailed your facility a couple of weeks ago.

You may recall that the letter introduced the Mammography Tumor Registry which provides follow-up mammography information to facilities and offers researchers the opportunity to gain insight into issues related to breast cancer detection. Thank you for promptly returning the reply card indicating how your facility collects and stores patient and exam information.

1. You indicated... *[restate information on card and make certain it is complete]. Photocopy of data collection card attached.*

Paper	**Computer**
_____ Patient Information	_____ Patient Information
_____ Health History	_____ Health History
_____ Exam Information	_____ Exam Information
_____ NO CHANGES	Which database system do you use?_____

2. We will be sending a packet of detailed information about the project, including forms that have to be signed, or checked for accuracy and returned to us. Who should receive this mailing? *(If the radiologist listed as responsible for MQSA QA, confirm spelling and address). (If someone else at facility, get spelling, title, and address).*

Name: _____
Address: _____
_____
_____

3. I want to be sure that we are making the appropriate contacts within your facility, so I'd like to ask if your clinic is associated with any other sites or locations?

1. *Yes:* a. _____
   b. _____
   c. _____
   d. _____

2. *No, this is the only location.* **(go to Q5)**

4. Does each site maintain a separate computer database of patient records, or are all records stored in one main database?

    1. *Each site has a separate data base.*
    2. *There is one main data base.*

    Comments:

    _____
    _____
    _____
    _____

5. Is there a primary business office?

    1. *Yes* **(go to Q6)**_____
    _____
    _____

    2. *No, this is the only office.* **(go to Q7)**

6. Who is the database manager in the business office? After the packet materials are returned to us, a data staff person from the project will be in touch with that contact person.

    Contact: _____
    _____
    _____
    _____

    Phone: _____

7. Who in your office is the database manager? After the packet materials are returned to use, a data staff person from the project will be in touch with that contact person.

    Contact: _____
    Phone: _____

Thank you for your time today. The packet of information will reach you within the next few weeks. We look forward to working with you.

Good-bye.

[redacted]

Form 3
Phone Contact Script
- Data Collection Method
Card Not Returned

Mammography Tumor Registry
Phone Contact A

Date: ___/___/___

Caller: _____

Facility Name _____

Address _____

Phone Number _____

Contact Person _____

Data Collection Card Returned? _____ yes *(after introduction, go to **section 1**)*
_____ no *(after introduction, go to **section 2**)*

Contact Notes:

_____
_____
_____
_____
_____
_____

INTRODUCTION

Hello. My name is [Marcia Gaul/Susie Wilson]. I'm calling from the Fred Hutchinson Cancer Research Center. Is the facility manager available to speak to me at this time?

1. *Yes:* Thank you.

2. *No: Arrange to leave a message or call back at a better time.*

Contact Name: _____
Title: _____
Phone Number: _____
Today's date: _____
Date to call back: _____

[redacted]

SECTION 2: ***Facilities that did not return the data collection card***

Hello. My name is [Marcia Gaul/Susie Wilson] and I'm calling from the FHCRC. I am phoning about the Mammography Tumor Registry, a research project being conducted here at the Center. This call is a follow-up to a letter we mailed your facility a couple of weeks ago.

1. Do you recall receiving a letter from us? A brochure and reply card were enclosed with the letter.

   1. *yes* **(go to Q3)**
   2. *no, confirm the address* **(then go to Q2)**

   *Address:*_____

   _____

2. The letter introduced a research study called the Mammography Tumor Registry which is being conducted by the Fred Hutchinson Cancer Research Center. In summary, the Mammography Tumor Registry provides follow-up mammography information to facilities and offers researchers the opportunity to gain insight into issues related to breast cancer detection. I would be happy to send you another copy of the letter and the informational brochure. Who should receive this mailing? After you've had the opportunity to read the information, I'll call you again and answer your questions and get some information.

   Name: _____
   Address: _____
   _____
   _____

3. I'm glad you received the mailing. I'd just like to remind you that the Mammography Tumor Registry is designed to provide follow-up mammography information to facilities and it offers researchers the opportunity to gain insight into issues related to breast cancer. To save you the trouble of filling out and returning the postcard, I can take that information over the phone. How does your facility collect and store patient information, health history, and exam information? *(check computer, paper, or both for each)*

	Paper	Computer
Patient Information	_____	_____
Health history	_____	_____
Exam information	_____	_____

Which database system do you use?_____

[redacted]

4. To make certain that we are contacting all of the facilities, I'd like to ask if your clinic has any other sites or locations?

   1. *Yes:* a. _____
      b. _____
      c. _____
      d. _____

   2. *No, this is the only location.* **(go to Q6)**

5. Does each site maintain a separate computer database of patient records, or are all records stored in one main database?

   1. *Each site has a separate data base.*
   2. *There is one main data base.*

   Comments:

   _____
   _____
   _____
   _____

6. Is there a primary business office?

   1. *Yes:* _____
      _____
      _____

   2. *No, this is the only office.* **(Go to Q8)**

7. Who in the business office is responsible for maintaining the patient information or database.

   Contact: _____
   Phone: _____

8. Who in your office maintains the patient information files or database? After the packet materials are returned to use, a data staff person from the project will be in touch with that contact person.

   Contact: _____
   Phone: _____

Thank you for your time today. You will receive the packet of information in the next few weeks. We look forward to working with you. Good-bye

Form 4
Facility - Data Release/ Confidentiality Agreement

MAMMOGRAPHY TUMOR REGISTRY

HUTCHINSON CANCER RESEARCH CENTER

DATA RELEASE/CONFIDENTIALITY AGREEMENT

The information exchanged between participating radiology facilities and the Mammography Tumor Registry (MTR) is of a highly sensitive nature, including confidential patient, radiologist, and cancer diagnosis data. It is absolutely essential to ensure, to the extent possible, that uses of MTR information is limited to internal improvement in the understanding of mammography. Uses for any other reason, particularly those resulting in personal disclosure, are strictly prohibited by law.

The MTR and the radiology facility named below mutually agree to the following:

1. The radiology facility agrees to provide patient-identified mammography data to the MTR for research purposes. MTR investigators and staff will take all reasonable precautions to ensure that the information provided to the MTR will remain confidential at all times. Fred Hutchinson Cancer Research Center agrees to use and disclose the information solely for research relating to the MTR. No facility-identified information will be disclosed for any reason without the consent of the participating radiology facility. All data contained in the MTR database is protected from disclosure or subpoena by a federal Certificate of Confidentiality issued under section 301(d) of the Public Health Service Act by the U.S. Department of Health and Human Services.

2. The radiology facility will take all reasonable precautions to ensure that all information provided by the MTR to the radiology facility in the form of feedback reports and/or data will remain confidential at all times. The facility agrees to use the information for quality improvement purposes only, including providing audit statistics to MQSA and/or the ACR, and agrees not to publish or disclose the information to any other person or entity. All information provided by the MTR will be considered a confidential medical record and is subject to Washington State law.

MTR

Nicole Urban, ScD Name	*Principal Investigator - MTR* Title
Nicole Urban Signature	*1100 Fairview Ave N, Seattle WA 98109* Address
9/17/97 Date	

Radiology Facility

Radiology Facility	Address
Name (print)	Title
Signature	Date

Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, MP-804, PO Box 19024, Seattle WA 98109-1024

[redacted]

Mammography Tumor Registry
Staff Directory

The following is the most current personnel information that we have for your facility. Please review this information, make corrections and additions as necessary, and return it in the envelope provided.

Facility Name:
Mailing Address:

Phone Number:

Key Contact Person: _____
Position: _____
Phone Number: (__) ___-____

Which staff member will provide the MTR with patient and exam data files or forms?
Example: work with MTR staff for initial preparation of data file, data documentation, etc.
Name: _____ **Phone Number:** (__) ___-____

Which staff member will review the data validation reports and make corrections?
Examples: different patient identification number assigned to same woman, discrepancies in assessment and follow-up codes, etc.
Name: _____ **Phone Number:** (__) ___-____

Who is to receive the Mammography Outcomes Report?
This report summarizes by assessment the mammography performed at your facility and the results of linking with the cancer registry.
Name: _____ **Position:** _____

Who is the radiologist responsible for MQSA audit analyses?
Name: _____ **Phone Number:** (__) ___-____

Mammography Technologist(s):

_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____
_____ ID: _____

[redacted]

Mammography Tumor Registry
Radiologist Directory

Physician(s):

Name:
Mailing Address:

Phone Number:

Name:
Mailing Address:

Phone Number:

Name:
Mailing Address:

Phone Number:

Name:
Mailing Address:

Phone Number:

Name:
Mailing Address:

Phone Number:

Name:
Mailing Address

Phone Number:

[redacted]

THE MAMMOGRAPHY TUMOR REGISTRY

Thank you for your interest in the Mammography Tumor Registry (MTR). Below are some questions and answers that will be helpful as you begin.

What is the Mammography Tumor Registry (MTR)?
The Mammography Tumor Registry is a research project that links patient and exam data provided by mammography facilities to cancer incidence data collected by the Washington State Cancer Registry and the Cancer Surveillance System. This combination of data enables the MTR to provide participating facilities with summary information regarding breast cancer rates among their mammography patients. In addition, these data create a rich pool of information for research investigation into breast cancer.

What kind of data will my facility be asked to provide to the MTR?
The data you contribute must include sufficient patient information to identify your patients in other databases. Specifically, this includes the patient's full name, birth date, Social Security number, and current address.

Since your data will be pooled with data from other mammography facilities, it must contain similar information to the data provided by those facilities. For example, each facility is asked to follow the ACR BIRADS assessment coding system of 0-5. If your facility does not use the ACR lexicon or uses a modified system, we will work with you to understand how your coding can be made compatible with the pooled data.

What staff responsibilities are incurred as we participate in the MTR?
Participation in the MTR requires some staff time and effort, but it should not impact your staff adversely. A staff person from the MTR will work with your facility staff person to familiarize you with the following:

- Submission of your database file to the MTR on a regular basis;
- Resolution of data discrepancies on the MTR validation form.

Our experience suggests that your initial contact is likely to take several hours over a week or two to prepare the initial data and go through the first validation report. Once the procedure becomes familiar, MTR related tasks should only take a few minutes a month. It is our hope that the patient follow-up information you receive from the MTR will reduce the amount of time you currently spend on patient follow-up.

How will MTR participation directly benefit my facility?
Following a data validation process and the linking of your mammography data with cancer incidence data, the MTR will send you a Mammography Outcomes Report. The report offers summary statistics of your facility's mammography results and any linked cancers, and provides the information needed to calculate positive predictive value,

THE MAMMOGRAPHY TUMOR REGISTRY

Thank you for your interest in the Mammography Tumor Registry (MTR). Below are some questions and answers that will be helpful as you begin.

What is the Mammography Tumor Registry (MTR)?

The Mammography Tumor Registry is a research project that links patient and exam data provided by mammography facilities to cancer incidence data collected by the Washington State Cancer Registry and the Cancer Surveillance System. This combination of data enables the MTR to provide participating facilities with summary information regarding breast cancer rates among their mammography patients. In addition, these data create a rich pool of information for research investigation into breast cancer.

What kind of data will my facility be asked to provide to the MTR?

The data you contribute must include sufficient patient information to identify your patients in other databases. Specifically, this includes the patient's full name, birth date, Social Security number, and current address.

Since your data will be pooled with data from other mammography facilities, it must contain similar information to the data provided by those facilities. For example, each facility is asked to follow the ACR BIRADS assessment coding system of 0-5. If your facility does not use the ACR lexicon or uses a modified system, we will work with you to understand how your coding can be made compatible with the pooled data.

As you begin participation, you are asked to consider the extent of the information you can provide. Because there is a growing trend in the research community to investigate risk factors and their relationship to early detection and disease progression, we will ask you to collect detailed health history information from your patients. This information may be more extensive than that which you currently collect.

How will participation in the MTR impact the facility staff?

Participation in the MTR requires some staff time and effort, but it should not impact your staff adversely. We first ask that you identify personnel who can assist with the following:

- Collection of detailed patient and exam information using a standardized form provided by the MTR;
- Routine submission of one copy of this data collection form to the MTR;
- Resolution of data discrepancies on the patient information and exam form.

Facility staff time will involve monthly submission of the patient information and exam forms to the MTR. Additional staff time may be required to check and resolve any data discrepancies we discover. Once the procedure becomes familiar, the MTR related tasks should only take a few minutes a month. It is our hope that the patient follow-

Form 8
Sample Validation Report
- Computer-based facilities

Mammography Tumor Registry
Data Validation Report

Facility: 000
Report Date: 4/21/97

Data Source: MRS Extract
Source Date: 12/01/96

Instructions

Please review the following information against your records.
Write corrections or other instructions after each item.
If an item is valid, check the "Valid as Shown" box for that item.
Please return the report in the envelope provided.

This report is ordered by patient account number.
Please note that items in the record description may also be problem items.

Record Description			Problem Description		Corrections
Acct / MR #	Name	Exam Date	Item Name	Item Value	Valid as Shown
xxxxxxxxx	xxxxxxxxxxxxx	04/15/96	Birth Date	02/15/1979	☐
xxxxxxxxx	xxxxxxxxxxxxx	02/01/94	Birth Date	04/01/1982	☐
xxxxxxxxx	xxxxxxxxxxxxx	07/15/96	City	Se	☐
xxxxxxxxx	xxxxxxxxxxxxx	04/01/96	Zip Code	9911	☐
xxxxxxxxx	xxxxxxxxxxxxx	01/12/96	State	RA	☐
xxxxxxxxx	xxxxxxxxxxxxx	09/10/96	Birth Date	03/01/1996	☐
xxxxxxxxx	xxxxxxxxxxxxx	12/11/96	City	Ke	☐
xxxxxxxxx	xxxxxxxxxxxxx	04/10/96	City	Re	☐
xxxxxxxxx	xxxxxxxxxxxxx	02/22/96	Zip Code	9811153	☐
xxxxxxxxx	xxxxxxxxxxxxx	08/04/96	State	RA	☐

Mammography Tumor Registry
Review of Assessments / Follow Up Recommendations

Facility: 000
Report Date: 4/21/97

Data Source: MRS Extract
Source Date: 12/01/96

Instructions

Please review the following information against your records.
Write corrections or other instructions on each line.
If the recommendation is valid, check the "Valid as Shown" box.
Please return the report in the envelope provided.

Please note that items describing a record (acct #, name, exam date) may also be included in the Data Validation Report as problem items.

This report is ordered by patient account number.

Acct / MR #	Name	Exam Date	Assessment	Recommendation(s)	Corrections	Valid as Shown
xxxxxxxxx	xxxxxxxxxxxxx	07/01/96	4 - Suspicious abnormality	Follow up mammogram		☐
xxxxxxxxx	xxxxxxxxxxxxx	06/15/96	4 - Suspicious abnormality	Follow up mammogram		☐
xxxxxxxxx	xxxxxxxxxxxxx	07/01/96	4 - Suspicious abnormality	Follow up mammogram		☐
xxxxxxxxx	xxxxxxxxxxxxx	04/12/96	5 - Highly suggestive of malig	None		☐
xxxxxxxxx	xxxxxxxxxxxxx	03/10/96	4 - Suspicious abnormality	Follow up mammogram		☐
xxxxxxxxx	xxxxxxxxxxxxx	03/11/96	1 - No rad evidence of malig	Surgical consult		☐
xxxxxxxxx	xxxxxxxxxxxxx	02/22/96	4 - Suspicious abnormality	Follow up mammogram		☐
xxxxxxxxx	xxxxxxxxxxxxx	08/04/96	0 - Incomplete - add'l eval	None		☐

Form 9
Patient History/Exam Form

PATIENT INFORMATION

Patient ID or file number:

Social Security Number: ___ ___ ___ - ___ ___ - ___ ___ ___ ___ Telephone Number: _____

First Name_____ Last Name_____ Middle Initial _____ Date of Birth ___/___/___

Street Address_____ City _____ State _____ Zip _____

ETHNIC BACKGROUND
- ☐1 Caucasian/White
- ☐2 African American/Black
- ☐3 Native American/Eskimo/Aleut
- ☐4 Asian/Pacific Islander
- ☐8 Other

HISPANIC/LATINA ORIGIN ☐0 No ☐1 Yes

EDUCATION *(check only one)*
- ☐1 1-11 Years
- ☐2 High school graduate
- ☐3 Some college/technical school
- ☐4 College graduate (4 years)
- ☐5 Post graduate degree

HEALTH INSURANCE *(check all that apply)*
- ☐1 None
- ☐1 Medicare
- ☐1 Medicaid
- ☐1 HMO, Managed Care
- ☐1 Private Insurance Company
- ☐1 Other
- ☐1 Not Sure

HISTORY

1. Have you ever had breast cancer?
☐0 No
☐1 Yes → If yes, which breast? ☐1 Left ☐2 Right ☐3 Both
Age at diagnosis? ___

2. Has your mother had breast cancer?
☐0 No
☐1 Yes → If yes, was she under age 50 when diagnosed? ☐0 No ☐1 Yes ☐9 Not sure
☐9 Not sure

3. How many of your sisters had breast cancer?
☐8 I have no sisters
☐0 None of my sisters
☐1 One sister
☐2 Two or more sisters
☐9 Not sure

If yes, were any of your sisters under age 50 when diagnosed?
☐0 No
☐1 Yes, one sister only
☐2 Yes, two or more sisters
☐9 Not sure

4. How many of your daughters had breast cancer?
☐8 I have no daughters
☐0 None of my daughters
☐1 One daughter
☐2 Two or more daughters
☐9 Not sure

If yes, were any of your daughters under age 50 when diagnosed?
☐0 No
☐1 Yes, one daughter only
☐2 Yes, two or more daughters
☐9 Not sure

5. Has any relative had ovarian cancer?
☐0 No
☐1 Mother, sister or daughter
☐2 Aunt or grandmother
☐3 Other relative
☐9 Not sure

6. Previous breast procedures *(check all that apply)*

	Left	Right	Both
Fine Needle Aspiration	☐1	☐2	☐3
Core Needle Biopsy	☐1	☐2	☐3
Open Excisional Biopsy	☐1	☐2	☐3
Lumpectomy	☐1	☐2	☐3
Mastectomy	☐1	☐2	☐3
Radiation Therapy	☐1	☐2	☐3
Reconstruction	☐1	☐2	☐3
Augmentation/Implants	☐1	☐2	☐3

7. Date of most recent breast biopsy:
___/___/___ ☐1 Never had a biopsy

8. Your age at the birth of your first child:
___ ☐00 I have no natural children

9. Have your menstrual periods stopped permanently? *(check only one)*
☐0 No
☐1 No, but my periods are less frequent
☐2 I now have bleeding from hormone replacement
☐3 Yes, my periods stopped naturally (menopause)
☐4 Yes, my periods stopped due to surgery
☐9 Not sure

If yes, how old were you when your periods stopped? ___

If no, what is the approximate length in days of your menstrual cycle? ___

And, what was the date of the start of your last menstrual cycle *(please estimate if you don't know the exact day)* ___/___/___

10. Have you had one or both ovaries removed?
☐0 No
☐1 Yes, one ovary removed
☐2 Yes, two ovaries removed
☐9 Not sure

11. Are you currently using any hormones? *(check all that apply)*
☐1 No
☐1 Yes, Estrogen only
☐1 Yes, Estrogen and Progesterone
☐1 Yes, Tamoxifen
☐1 Yes, birth control
☐1 Yes, other hormone
☐1 Not sure

12. Have you had any problems or symptoms with your breasts in the last 3 months?
☐0 No ☐1 Yes

If yes, check all that apply:

	Left	Right	Both
Lump	☐1	☐2	☐3
Nipple discharge	☐1	☐2	☐3
Pain	☐1	☐2	☐3
Skin changes	☐1	☐2	☐3
Other	☐1	☐2	☐3

13. Did you make this appointment due to a concern about a breast problem found in the past 3 months? *(check one)*
☐0 No, this is a routine mammogram.
☐1 Yes, I found something new.
☐2 Yes, my doctor found something new.
☐3 Yes, I have a general concern but no specific symptoms.

14. Have you had a previous mammogram?
☐0 No ☐1 Yes ☐9 Not sure

If yes, what was the date of your last mammogram? ___/___/___

15. Have you ever had a clinical breast exam (a physical breast exam performed by a health care provider)?
☐0 No ☐1 Yes ☐9 Not sure

If yes, how long since your last clinical breast exam? *(Check only one)*
☐1 Within the last 3 months
☐2 3 to 12 months
☐3 More than 1 year ago

DO NOT WRITE BELOW THIS LINE

EXAM INFORMATION (To be completed by clinic personnel)

Facility/Exposure Site _____ Tech ID _____ Radiologist ID _____ Date of Mammogram ___/___/___

1. Physical exam results
☐0 Negative
☐1 Positive (suspicious for malignancy)
☐2 Not performed

2. Symptoms *(check all that apply)*
☐1 None
☐1 Lump
☐1 Bloody nipple discharge
☐1 Pain
☐1 Other: _____

3. Was patient referred because of symptoms detected by CBE performed within the last 3 months?
☐0 No ☐1 Yes ☐9 Unknown

4. Date of last mammogram ___/___/___

5. Comparison films available? ☐0 No ☐1 Yes

6. Reason for mammogram *(check only one)*
☐0 Screening (asymptomatic)
☐1 Diagnostic (symptomatic)
☐2 Short interval follow-up
☐3 Additional view(s) for current exam
☐4 Special study
☐8 Other: _____

7. Procedure
☐3 Bilateral mammography
☐2 Right only
☐1 Left only

8. Density (code breast with greatest density)
☐1 Mostly fatty
☐2 Scattered fibroglandular tissue
☐3 Heterogeneously dense
☐4 Extremely dense

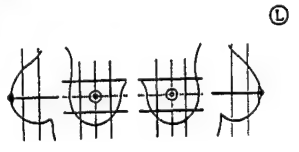
9. Assessment - Right Breast
☐0 Needs additional evaluation
☐1 Normal
☐2 Benign finding
☐3 Probably benign; short follow-up
☐4 Suspicious abnormality
☐5 Highly suspicious for malignancy

10. Assessment - Left Breast
☐0 Needs additional evaluation
☐1 Normal
☐2 Benign finding
☐3 Probably benign; short follow-up
☐4 Suspicious abnormality
☐5 Highly suspicious for malignancy

11. Assessment based on: *(check all that apply)*
☐1 Basic 2 views per breast
☐1 Additional views
☐1 Left breast ☐2 Right breast ☐3 Both breasts
☐1 Clinical findings
☐1 Referring physician's report
☐1 Comparison with previous films
☐1 Patient report
☐1 Ultrasound
☐1 Family history
☐1 Patient history

12. Recommendation for mammogram follow-up
☐1 Routine follow up interval Months: _____
☐1 Short term follow up Months: _____

13. Recommendation for immediate work-up *(check all that apply)*
☐1 Additional views
☐1 Ultrasound
☐1 FNA
☐1 Core needle biopsy
☐1 Surgical biopsy
☐1 MRI
☐1 Surgical or clinical consult
☐1 Other immediate workup: _____



Signature

○ = lump ✳ = scar
● = mole × = pain

MAMMOGRAPHY QUALITY IMPROVEMENT PROJECT

FRED HUTCHINSON CANCER RESEARCH CENTER

RADIOLOGIST'S PARTICIPATION AGREEMENT

INVESTIGATORS:

Nicole Urban, ScD, Principal Investigator, (206) 667-6560
Mariann Drucker, MD, Radiologist, (206) 667-6560
Connie Lehman, MD, PhD, Radiologist, (206) 543-3320

ALL WORK CONDUCTED BY:

Fred Hutchinson Cancer Research Center
Cancer Prevention Research Program
1100 Fairview Ave. N (MP 804)
PO Box 19024
Seattle, WA 98109-1024

BACKGROUND:

Breast cancer is one of the major causes of death among women age 50 or older. At least one in nine women in the United States will be diagnosed with breast cancer in her lifetime. Despite research efforts in breast cancer prevention, a woman's best hope for cancer survival lies in identification of the disease early enough to treat it effectively. The primary tool for early detection is regular screening mammography. In order to improve the quality of screening mammography, it is hypothesized that a Continuous Quality Improvement (CQI) program, which provides radiologists with immediate feedback on their interpretation of a set of mammograms selected for their teaching value, can significantly improve the ability of the radiologist to detect early stage breast cancer.

PURPOSE:

We are asking you to consent to participate in a randomized study being conducted within a research project, the **Mammography Quality Improvement Project (MQIP).** The **primary research objective** is to determine if a Continuous Quality Improvement (CQI) program can increase the accuracy of mammography film interpretation by radiologists. The **secondary research objectives** are to 1) determine inter-rater variability in film interpretation, before and after implementation of the CQI program, in a set of mammograms selected for their teaching value; 2) determine whether digitized films can be interpreted with the same accuracy as high-quality copies of films; and 3) determine if the accuracy of film interpretation depends on covariates such as the woman's age and breast density.

PROCEDURES:

You were selected for participation in this study because you regularly read mammograms for a Washington State mammography facility that is part of the Mammography Tumor Registry. These facilities will be rank ordered by the number of radiologists reading mammograms for each, and then paired. Randomization to the CQI program group or to the control group will take place within these pairs.

Participant's Initials/Date

Accordingly, you have a 50% chance of being randomized to participate in the CQI program. The CQI program procedures apply to you **only** if you are randomized to the film reading study.

The CQI program consists of five sessions and involves reading a total of 180 mammography films (cases) randomly assigned to five sets – 2 sets of 45 films and 3 sets of 30 films. Radiologists are randomly assigned sets of films to be read at each of the sessions. Scheduled over a 12 – 14 month period, the five two hour film-reading sessions will be conducted individually, on-site, and will be facilitated by a project field coordinator. Sessions 1 – 4 will utilize both films and digitized images. Session 5 offers only digitized images in order to determine the feasibility of CQI program dissemination via the Internet.

For each session, an appointment will be a made for you to read a set of films and code your assessment of each case on a laptop computer. The films will be high-quality copies, because the originals are not available to the project for the requisite 18 month period. You will be instructed to read the films as you generally read screening mammograms with one exception – the cancer prevalence in the population from which the films were selected is higher than you would expect in your practice.

Your assessment will be made using the standard ACR 5-point scale as follows:

- 1 indicates "normal, routine follow-up recommended,"
- 2 indicates "benign, routine follow-up,"
- 3 indicates "probably benign, early recall recommended,"
- 4 indicates "suspicious for cancer, consider biopsy,"
- 5 indicates "highly suspicious for cancer, biopsy recommended."

You will also be asked to indicate the suspicious abnormality on the digitized computer image.

You must read and assess all films in the set before the computer offers feedback. This feedback consists of the correct interpretation of each film with indication of the areas of benign or malignant disease. Text will explain the nature of disease (if any) and the strategy for interpretation.

FILM SOLICITATION:

The 180 films used in the study will include women with and without cancer, as well as both younger and older women. We anticipate that up to one-half of the films will be contributed by the participating radiologists. As a participant, you are asked to submit interesting films for possible inclusion in the study. The mammograms must be from women who were asymptomatic at the time of the film. At least one of the films should be from a woman <50 years of age. Please submit films that were challenging to interpret. Malignant cases must have biopsy correlation. Benign assessment is determined either by a) a woman's cancer-free status two years after the mammogram; b) additional evaluation resulting in a benign finding; or c) negative biopsy results.

Other films used in the sessions will be selected from radiologists around the state and will be supplemented with interesting cases from the teaching files of the radiologist investigators. Confidentiality regarding the source of the film is maintained at all times. Before a film is reviewed for inclusion in the study, it will be masked. There will be no patient, radiologist, or facility identifiers on any of the study films.

CONTINUING MEDICAL EDUCATION:

The CQI program will be an ACR/UW accredited Continuing Medical Education (CME) program. It is anticipated that a 2 hour appointment will be required for each film reading session. Two CME credits will be awarded for each session, for a total of 10 CME hours.

RISK:

The only possible risk involved in study participation is the potential loss of privacy associated with the data describing performance on the mammography practice readings. To minimize this risk, each radiologist is assigned a personal identification number. In the data files, you are identified only by this

number. The information linking names with numbers will be stored in sealed files at the Fred Hutchinson Cancer Research Center. Once the necessary information is obtained from data files, all identifying information will be destroyed. Personal and facility identity will not be revealed in any publication or in the results of the study. Study records will be maintained indefinitely for the purpose of analysis and are accessible only by research staff and the US Army Medical Research and Materiel Command.

Representatives from the U.S. Army Medical Research and Materiel Command (and, where applicable, the Food and Drug Administration, and the U.S. Army Medical Department Center and School) may inspect the records of the research in their duty to protect human subjects in research.

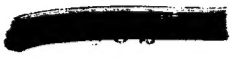
OTHER PERTINENT INFORMATION:

Participation is completely voluntary. If you agree to participate, you are free to withdraw at any time. Your refusal to participate or a decision to discontinue participation will involve no penalty or loss of benefits to which you are otherwise entitled.

The Department of Defense is funding this research project. Should you be injured as a direct result of participating in this research project you will be provided medical care, at no cost to you, for that injury. You will not receive any injury compensation, only medical care. You should also understand that this is not a waiver or release of your legal rights. You should discuss this issue thoroughly with the Principal Investigator before you enroll in this study. There is no compensation for involvement in this study.

For further information about the study please contact the project office at (206) 667-6560 or (800) 224-2331. If you have any questions about your rights as a study participant, please contact Karen Hansen in the Institutional Review Office of the Fred Hutchinson Cancer Research Center at (206) 667-4867.

Participant's Initials/Date



INVESTIGATOR'S STATEMENT

I have provided an explanation of the above research program. The subject was given an opportunity to discuss the procedures, including possible alternatives, and to ask additional questions. A signed copy of the consent form has been given to the subject.

INVESTIGATOR SIGNATURE DATE

SUBJECT'S STATEMENT

I agree to participate in this study (and to the conditions outlined in this consent form that I have read and signed). I had the opportunity to ask questions about the study and my participation. All questions were answered to my satisfaction. I understand that future questions I may have about the research will be answered by the project investigators listed herein and that any questions I have about my rights as a research subject will be answered by them. I give permission for my study records to be available to physicians, scientists, and personnel working on this study at the Fred Hutchinson Cancer Research Center. Records are also available to the US Army Medical Research and Materiel Command. I acknowledge that I will receive a signed copy of this consent form. I read and initialed each page of the form.

PARTICIPANT SIGNATURE DATE

PRINTED NAME

ADDRESS

CITY, STATE, ZIP

WITNESS SIGNATURE

PRINTED NAME

ADDRESS

CITY, STATE, ZIP

For MQIP Office Use Only
MQIP ID # _____

Form 12
Case Selection Form

MAMMOGRAPHY QUALITY IMPROVEMENT PROGRAM

Instructions for Case Selection

Mammography Tumor Registry
Fred Hutchinson Cancer Research Center
1100 Fairview Ave N (MP-804)
PO Box 19024
Seattle, WA 98109-1024

Please choose mammograms from asymptomatic women that offered you a diagnostic challenge. Include an equal number of benign and malignant assessments. Include at least one case from a woman under 50. Other criteria include:

- The film is of educational interest because interpretation was challenging. Malignant cases must have biopsy correlation. Benign assessment is confirmed by a) a woman's cancer-free status two years after the mammogram; b) additional evaluation with a benign finding; or c) negative biopsy results.
- Submit cases with only one challenging finding; exclude cases with either two synchronous cancer sites in separate quadrants of one breast or cases with bilateral breast cancer.
- Send no mammograms at this time. Fill out this form completely and the original films will be requested from your film room. Films will be masked for identifiers, copied, and then selected. Original films will be returned within eight weeks. For emergency immediate film return, call (800) 224-2331 or (206) 667-6560.

Patient Information

Patient's name: _____ **Medical ID#:** _____

Prior name(s) used by the patient: _____ Date of Birth: ___/___/___

Social Security Number: ___ ___ ___ - ___ ___ - ___ ___ ___ ___

Case Information

Date(s) of the mammogram(s) to be considered: ___/___/___

Facility where mammogram is stored: _____

Areas of interest/Description of case: _____

Classification of finding(s): *Check all that apply.*

- ☐ Benign/Normal Mammogram
- ☐ Calcifications
- ☐ Masses
- ☐ Architectural distortion

ACR Assessment

- ☐ 1 Normal
- ☐ 2 Benign finding
- ☐ 3 Probably benign; short follow-up
- ☐ 4 Suspicious abnormality
- ☐ 5 Highly suspicious for malignancy

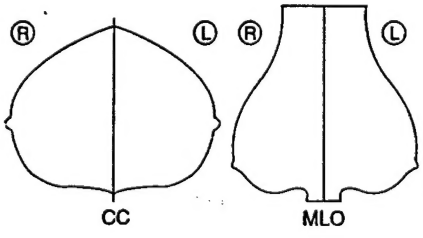
Date of biopsy: (if performed) ___/___/___

Biopsy results:

☐ Benign
- ☐ Normal ductal epithelium
- ☐ Fibroadenoma
- ☐ Cyst
- ☐ Sclerosing Adenosis
- ☐ "Fibrocystic changes"
- ☐ Other _____

☐ Malignant
- ☐ DCIS
- ☐ Invasive ductal carcinoma
- ☐ LCIS
- ☐ Invasive lobular carcinoma
- ☐ Other _____

Location of finding:



I authorize temporary release of these films to the Fred Hutchinson Cancer Research Center for educational purposes.

_____
Physician's Name (please print)

_____
[redacted]hysician's Signature

caseselect